AD_____

Award Number: W81XWH-05-1-0349

TITLE: Molecular Basis of Essential Thrombocytosis

PRINCIPAL INVESTIGATOR: Wadie F. Bahou, M.D.

CONTRACTING ORGANIZATION: The Research Foundation of
SUNY State University of New York
Stony Brook, NY 11794-8151

REPORT DATE: June 2008

TYPE OF REPORT: Final

PREPARED FOR: U.S. Army Medical Research and Materiel Command
Fort Detrick, Maryland 21702-5012

DISTRIBUTION STATEMENT: Approved for Public Release;
Distribution Unlimited

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. REPORT DATE	2. REPORT TYPE	3. DATES COVERED
1 Jun 2008	Final	18 May 2005 – 17 May 2008

4. TITLE AND SUBTITLE

Molecular Basis of Essential Thrombocytosis

5a. CONTRACT NUMBER

5b. GRANT NUMBER
W81XWH-05-1-0349

5c. PROGRAM ELEMENT NUMBER

6. AUTHOR(S)

Wadie F. Bahou, M.D.

E-Mail: wbahou@notes.cc.sunysb.edu

5d. PROJECT NUMBER

5e. TASK NUMBER

5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)

The Research Foundation of SUNY
State University of New York
Stony Brook, NY 11794-8151

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)
U.S. Army Medical Research and Materiel Command
Fort Detrick, Maryland 21702-5012

10. SPONSOR/MONITOR'S ACRONYM(S)

11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION / AVAILABILITY STATEMENT
Approved for Public Release; Distribution Unlimited

13. SUPPLEMENTARY NOTES

14. ABSTRACT

Human blood platelets play critical roles in normal hemostatic processes and pathologic conditions such as thrombosis (i.e. cardiovascular disease and stroke), vascular remodeling, inflammation, and wound repair. Despite the biological importance of platelets and their intact protein synthetic capabilities, remarkably little is known about platelet mRNAs. The pathogenesis of essential thrombocytosis (ET), a disease of platelet number and function, is poorly understood at the molecular level. The main goal of this project is to build on our preliminary data that suggests that patients with ET have distinct platelet transcript profiles that differ from those of normal platelets. The three main hypotheses to be tested are: (1) patients with ET have mRNA profiles that are distinct from those of normal controls; (2) these differences can be used to elucidate the molecular basis of ET; and (3) these differences can be used to differentiate ET from other causes of thrombocytosis (ET diagnostics). Completion of the specific aims as outlined below should (i) provide considerable insight into the molecular basis of ET, (ii) assist with molecular diagnostics, and (iii) help to devise rational approaches for pharmacological intervention.

15. SUBJECT TERMS
Myeloproliferative disorders, thrombocytosis

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT	b. ABSTRACT	c. THIS PAGE			19b. TELEPHONE NUMBER *(include area code)*
U	U	U	UU		

Table of Contents

I. INTRODUCTION. Human blood platelets play critical roles in normal hemostatic processes and pathologic conditions such as thrombosis (i.e. cardiovascular disease and stroke), vascular remodeling, inflammation, and wound repair. Despite the biological importance of platelets and their intact protein synthetic capabilities, remarkably little is known about platelet mRNAs. The pathogenesis of essential thrombocytosis (ET), a disease of platelet number and function, is poorly understood at the molecular level. The main goal of this project is to build on our preliminary data that suggests that patients with ET have distinct platelet transcript profiles that differ from those of normal platelets. The three main hypotheses to be tested are: (1) patients with ET have mRNA profiles that are distinct from those of normal controls; (2) these differences can be used to elucidate the molecular basis of ET; and (3) these differences can be used to differentiate ET from other causes of thrombocytosis (ET diagnostics). Completion of the specific aims as outlined below should (*i*) provide considerable insight into the molecular basis of ET, (*ii*) assist with molecular diagnostics, and (*iii*) help to devise rational approaches for pharmacological intervention. These three major hypotheses are the basis for the following specific aims:

Specific Aim I. Validate differences in microarray profiles between normal and ET patients

Specific Aim II. Establish the positive and negative predictive values of these profiles

Specific Aim III. Develop a pilot ET diagnostic microarray chip and evaluate a transcriptome-based approach to the diagnostics of ET.

II. BODY.

A. SPECIFIC AIMS I AND II.

1. APPROACH: Microarray analysis was used to study the molecular basis of essential thrombocythemia using highly-purified platelets isolated from peripheral blood (20 mL) or by apheresis. Differences were validated by quantitative PCR (Q-PCR) and/or protein analyses as necessary.

2. RESULTS: Initial one-way ANOVA identified 170 genes that were differentially expressed, the majority of which (141) were up-regulated in ET platelets; only 29 genes were down-regulated in ET compared to normal platelets [*Ref. 1; Figure 2*]. Analysis of the smaller subset of platelet-restricted genes demonstrated that only 13 genes were differentially-expressed (12 up-regulated, 1 down-regulated) in ET. A disproportionate number of upregulated genes encoded proteases or protease inhibitors (*HPSE*, *MMP1*, *SERPINI1*), a class of proteins known to be associated with tumor invasiveness and metastases. The single down-regulated gene (*HSD17B3*) was present in all normal samples and absent in all ET platelets [*Ref. 1, Figure 3*]. *HSD17B3* belongs to an extended family of 17BHSDs retaining oxido-reductase activity toward discrete substrates, and encodes an enzyme (type 3 17β-hydroxysteroid dehydrogenase) previously described as testis-specific. This enzyme is known to catalyze the penultimate step in testosterone biosynthesis. Using a functional assay of testosterone generation, we demonstrated that platelets retained 17βHSD3 activity, with nearly 10% of the capacity found in mouse testis, providing evidence for the first non-testicular source of this enzyme [*Ref. 1, Figure 6*]. Transcripts for two additional members of this family were found in human platelets, one of which (*HSD17B12*) was upregulated in the initial cohort of patients studied. Subsequent Q-PCR results were entirely concordant for all individuals studied, demonstrating that *17BHSD12:17BHSD3* transcript ratios reliably distinguished ET from normal patient platelets in all samples studied to date (N=20; 6 apheresis samples, 14 peripheral blood samples; $p < 0.0001$; [*Ref. 1; Figure 5*]). Furthermore, these differential patterns of *HSD17B* expression appeared unrelated to the development of thrombocytosis *per se*, but rather, were restricted to the ET phenotype.

B. SPECIFIC AIM III: Development of a diagnostic platelet oligonucleotide microarray chip

1. RESULTS (Chip design and fabrication). A well-characterized 70-mer oligonucleotide gene set was used as the starting point for our gene chip. The Qiagen probe set contains 34,580 optimized 70-mer oligonucleotides, representing 24,650 genes and 37,123 transcripts. The design is based on the Ensemble (www.ensembl.org/) Human 13.31 Database and the Human Genome sequencing project, and directly deals with alternative splicing variants using common, partial common, or transcript oligonucleotides. The oligonucleotide probe set comes with complete annotation including Gene Ontology, oligonucleotide chromosome coordinates, comparative genomic analysis, and other functional annotations (i.e InterPro, OMIM, etc.). For our purposes, the final list was generated by data-mining of the original 14-chip analyses (*refer to* **Fig. 1**), and incorporated the following gene cohorts: (*i*) a group of platelet-restricted genes with no expression in leukocytes; (*ii*) the preliminary group of discriminatory genes distinguishing between thrombohemorrhagic ET phenotypes; (*iii*) the list of genes with platelet expression > leukocyte expression by 10-fold; and (*iv*) the list of genes with leukocyte expression > platelet expression by 10-fold [additional leukocyte "contamination" control]. Further, we have now analyzed the relative gene expression of the platelet list and found that the expression patterns logically follow a Poisson distribution. To date we have printed 100 slides and quality-controlled the first ten with excellent reproducibility in spotting (**Fig. 1**). After removal of duplicates, the final list contains 432 genes which clearly co-segregate by cell-type as demonstrated *in silico [Ref. 4; Figure 3*]. The chip may be readily modified as more informative genes are identified.

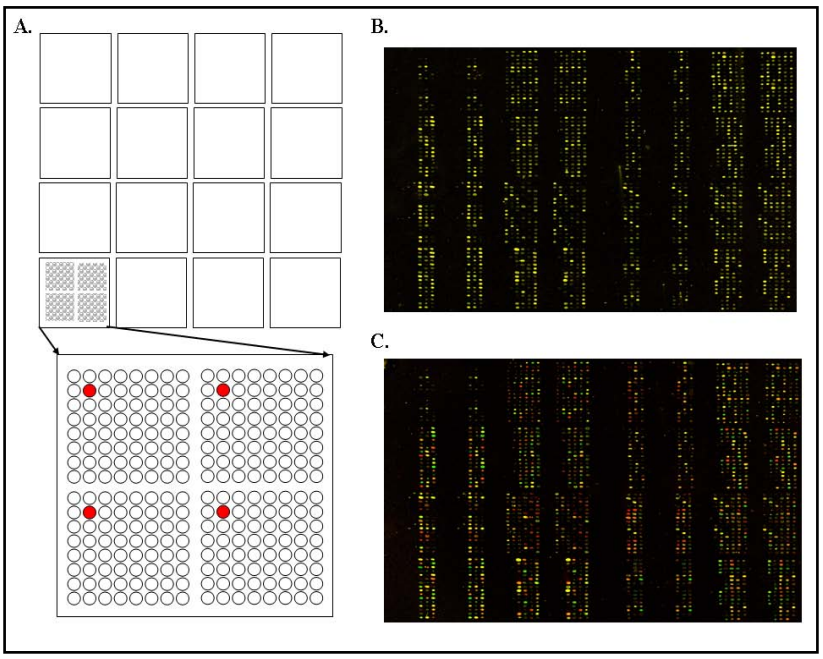


Fig. 1. Oligonucleotides (ODNs) representing the 432-member gene list (along with positive [ribosomal RNAs, housekeeping genes, etc.) and negative [i.e. *Arabidopsis*] controls) were synthesized as 70-mers and used for spotting onto glass slides[1]. ODNs are each represented 4 times/slide (represented as red dots in *Panel A*) as a means of obtaining reliable, integrated expression median values; the final chip contains 2,132 oligonucleotides. For these experiments, two samples are differentially labeled (i.e. either Cy5- or Cy3-), followed by hybridization onto the gene chip. Note that co-hybridization of Cy5- or Cy3-labeled Universal reference RNA (Strategene) provides clear evidence for successful spotting, labeling, and hybridization as demonstrated by the yellow signals from the majority (if not all) of the spotted oligonucleotides (*Panel B*). When we repeated these experiments using platelet (Cy5):human reference RNAs (Cy3), completely different patterns were evident (*Panel C*), all as predicted.

a. Application of the novel platform to ET: Development of class prediction models for thrombocytosis. Hematologic criteria for distinguishing among the various causes of thrombocytosis are limited in their capacity to delineate clonal (essential thrombocythemia; ET) from non-clonal (reactive thrombocytosis; RT) etiologies. We have developed algorithms that can predict phenotypic class using distinct genetic biomarker subsets. Using the focused microarray gene chip, we studied the platelet genetic profiles of 95 patients (48 healthy controls, 23 RT, 24 ET [11 contained the $Jak2V^{617}F$ mutant allele; 13 were homozygous normal]). Gender differences in expression profiles were analyzed by cohort. Linear discriminant analysis with cross-validation was used to identify subsets that segregated phenotypes based on microarray profiles. Cross-platform consistency of platelet gene profiles was validated using quantitative real-time reverse-transcription polymerase chain reaction (qRT-PCR). Class prediction algorithms using the genetic biomarkers were developed to assign phenotypic class between the thrombocytosis cohorts, and by Jak2 genotype (within the ET cohort).

Gender differences were rare in normal and ET cohorts (<1% of genes), but were male-skewed for ~3% of RT genes. An 11-biomarker gene subset was identified that discriminated among the three cohorts with 86.3% accuracy [*Ref. 11*, *Figure 1*]. Two-way class prediction (ET *vs.* RT) was successful in 93.6% of patients using

microarray profiling and 90.0% of patients using qRT-PCR [*Ref. 11; Figure 2*]. A 4-biomarker gene subset predicted Jak2-wild type ET with >90% accuracy using either microarray or qRT-PCR profiling, with less robust class prediction in Jak2V^{617}F-containing ET patients. Based on these studies, we established that distinct genetic biomarker subsets can be developed to predict thrombocytosis class and Jak2V^{617}F-negative ET using routine phlebotomy [*Ref. 11*, *in revision].*

III. KEY RESEARCH ACCOMPLISHMENTS.

- First demonstration that platelet microarray studies can be applied to a human platelet disorder
- First demonstration that ET platelets have a profile distinct from normal platelets
- First demonstration that a diagnostic assay may be available to distinguish ET from normal platelets
- Initial fabrication, characterization, and validation of a platelet-diagnostic oligonucleotide chip

IV. REPORTABLE OUTCOMES (2005-present).

A. Manuscripts accepted and/or published

- [1] Gnatenko, D., L. Cupit, E. Huang, A. Dhundale, P. Perrotta, and W. Bahou. Platelets express steroidogenic 17β-hydroxysteroid dehydrogenases: distinct profiles predict the essential thrombocythemic phenotype, Thromb. Haemost., 94:412-21, 2005.

- [2)] Watson, S., W. Bahou, D. Fitzgerald, W. Ouwehand, A. K. Rao, and A. Leavitt. Mapping the platelet proteome: a report of the ISTH platelet physiology subcommittee. J. Thr. Haemost. 3:2098-2101, 2005.

- [3] Gnatenko, D. and W. Bahou. Recent advances in platelet transcriptomics. Trans. Med. Hemotherapy
33:217-226, 2006

- [4] Gnatenko, D., P. Perrotta, and W. Bahou. Proteomic approaches to dissect platelet function: half the story. Blood 108:3983-3991, 2006.

- [5] Bahou, W. Megaprofiles provide big insights into platelet function. Blood 109:3120-3121, 2007.

- [6)] Senzel, L., D. Gnatenko, and W. Bahou. Patelet transcriptome and cardiovascular disease. Future Cardiology 3:391-398, 2007.

- [7] Gnatenko, D., J. Dunn, J. Schwedes, and W. Bahou. Transcript profiling of human platelets using microarray and SAGE. Methods Molec. Biol. 496:247-274, 2008.

- [8] Senzel, L, D. Gnatenko, and W. Bahou**.** Genomic and proteomic applications in platelet diagnosis and classification. Sem. Thromb. Haemostasis, *in press*.

- [9] Senzel, L, D. Gnatenko, and W. Bahou**.** Proteomics and transcriptomics in the development of antithrombotics, *in* New Therapeutic Agents in Thrombosis and Thrombolysis*,* Freedman and Loscalzo*,* eds*. in press.*

- [10] Gnatenko, D., W. Zhu, and W. Bahou. Multiplexed genetic profiling of human blood platelets using fluorescent microspheres, Thromb. Haemostasis 100:929-936, 2008.

B. Manuscripts In Revision

- [11] Gnatenko, D., W. Zhu, X. Xu, M. Zarrabi, C. Kim, A. Dhundale, and W. Bahou. Molecular classification and prediction models of thrombocytosis, NEJM.

C. Newly funded awards

R21HL086376 (Bahou) Project period: 9/1/06 - 8/31/08
NIH/NHLBI
Integrated molecular profiling of the thrombohemorrhagic phenotype

D. Pending awards

"*Genetic dissection of the platelet thrombohemorrhagic phenotype*" (NIH/NHLBI)

PRIORITY SCORE 160 (18.4%)

"*Platelet steroidogenic pathway genes and stroke susceptibility in Women's Health Initiative (WHI) cohorts*"
(NIH/NHLBI)

V. CONCLUSION: We have made considerable progress in all aims of the research, specifically in (*i*) providing the proof-of-principal that ET has unique mRNA transcript profiles, (*ii*) that these profiles can be used to develop potential diagnostic tests for the disease, and (*iii*) in identifying a novel steroidogenic pathway in normal and platelets. Furthermore, we have now created a unique platelet microarray chip that is being developed for more robust profiling. As importantly, we have now applied this unique infrastructure to develop class prediction models for thrombocytosis that have broad clinical relevance, as highlighted by a favorably reviewed manuscript currently under review by NEJM [*Ref. 11*]. This research has opened up new avenues for research, and provided our group with additional sources of extramural funding to continue research in this area. With subsequent funding opportunities, we anticipate developing diagnostic biomarkers not only for thrombocytosis, but also for thrombotic or hemorrhagic risk known to be associated with ET.

VI. REFERENCES. SEE SECTION IV. Reportable Outcomes (*above*)

VII. REFERENCED AND SELECTED APPENDICES (updated since last progress report).

- [1] Gnatenko, D., L. Cupit, E. Huang, A. Dhundale, P. Perrotta, and W. Bahou. Platelets express steroidogenic 17β-hydroxysteroid dehydrogenases: distinct profiles predict the essential thrombocythemic phenotype, Thromb. Haemost., 94:412-21, 2005

- [4] Gnatenko, D., P. Perrotta, and W. Bahou. Proteomic approaches to dissect platelet function: half the story. Blood 108:3983-3991, 2006.

- [7] Gnatenko, D., J. Dunn, J. Schwedes, and W. Bahou. Transcript profiling of human platelets using microarray and SAGE. Methods Molec. Biol. 496:247-274, 2008.

- [10] Gnatenko, D., W. Zhu, and W. Bahou. Multiplexed genetic profiling of human blood platelets using fluorescent microspheres, Thromb. Haemostasis 100:929-936, 2008.

- [11] Gnatenko, D., W. Zhu, X. Xu, M. Zarrabi, C. Kim, A. Dhundale, and W. Bahou. Molecular classification and prediction models of thrombocytosis, NEJM.

VIII. SUPPORTING DATA. As outlined in Figure 1 (*above*) and in the appended articles, the platelet oligonucleotide chip provides the requisite tool to screen large populations to optimally establish normal profiles and delineate any inter-racial variability (if it exists). Methods to differentiation ET from RT have been developed and validated using both PCR-based technology for select genes, and the oligonucleotide chip. The clinical relevance of these observations is demonstrated by the ongoing review of this manuscript by a clinically visible journal [*Ref. 11*].

IX. INDIVIDUALS RECEIVING PAY

Jean Wainer, BS Research Support Specialist

Platelets express steroidogenic 17β–hydroxysteroid dehydrogenases

Distinct profiles predict the essential thrombocythemic phenotype

Dmitri V. Gnatenko[1], Lisa D. Cupit[1], Emily C. Huang[2], Anilkumar Dhundale[2], Peter L. Perrotta[3], Wadie F. Bahou[1,4]

[1]Department of Medicine, State University of New York, Stony Brook, New York, USA
[2]Center for Biotechnology and Department of Bioengineering, State University of New York, Stony Brook, New York, USA
[3]Department of Pathology, State University of New York at Stony Brook, New York, USA
[4]Program in Genetics, State University of New York at Stony Brook, New York, USA

Summary

Human blood platelets have important, regulatory functions in diverse hemostatic and pathological disorders, including vascular remodeling, inflammation, and wound repair. Microarray analysis was used to study the molecular basis of essential thrombocythemia, a myeloproliferative disorder with quantitative and qualitative platelet defects associated with cardiovascular and thrombohemorrhagic symptoms, not infrequently neurological. A platelet-expressed gene (*HSD17B3*) encoding type 3 17β-hydroxysteroid dehydrogenase (previously characterized as a testis-specific enzyme catalyzing the final step in gonadal synthesis of testosterone) was selectively down-regulated in ET platelets, with reciprocal induction of the type 12 enzyme (*HSD17B12*). Functional 17β-HSD3 activity corresponding to ~10% of that found in murine testis was demonstrated in normal platelets. The induction of *HSD17B12* in ET platelets was unassociated with a concomitant increase in androgen biosynthesis, suggesting distinct functions and/or substrate specificities of the types 3 and 12 enzymes. Application of a molecular assay distinguished ET from normal platelets in 20 consecutive patients ($p < 0.0001$). These data provide the first evidence that distinct subtypes of steroidogenic 17β–HSDs are functionally present in human blood platelets, and that the expression patterns of *HSD17B3* and *HSD17B12* are associated with an uncommon platelet disorder manifest by quantitative and qualitative platelet defects.

Keywords

Platelets, hydroxysteroid dehydrogenases, microarray, essential thrombocythemia

Thromb Haemost 2005; 94: 412-21

Introduction

Circulating blood platelets are anucleate although they retain small amounts of megakaryocyte-derived mRNAs and a fully functional protein biosynthetic capacity (1). Essential thrombocythemia (ET) represents a myeloproliferative disorder subtype, characterized by increased proliferation of megakaryocytes, elevated numbers of circulating platelets, and considerable thrombohemorrhagic events, not infrequently neurological (2). ET is seen with equal frequency in males and females, although an additional female incidence peak at age 30 may explain the apparent higher disease prevalence in females. The molecular basis of ET remains unestablished although historically it has been considered a "clonal" disorder (3). Causative mutations have been identified in the thrombopoietin gene, but these appear to be uncommon, and restricted to rare individuals with familial thrombocythemia (4). Other than the exaggerated platelet volume and inconsistent platelet aggregation abnormalities evident in subsets of ET platelets, no functional or diagnostic test is currently available for ET and it remains a diagnosis of exclusion (5).

The utility of gene expression profiling for the molecular classification of human cancer is well-documented, although its applicability to poorly-understood myeloproliferative disorders such as essential thrombocythemia remains unestablished. Previous work from this laboratory demonstrated the feasibility of

Correspondence to:
Wadie F. Bahou
Division of Hematology
HSCT15–040
State University of New York at Stony Brook
11794–8151 New York
New York, USA
Tel: + 1 631–444–2059, Fax: + 1 631–444–7530
E-mail: wbahou@notes.cc.sunysb.edu

Received January 18, 2005
Acceptzed after resubmission May 29, 2005

Financial support:
This work was supported by grants HL49141, HL76457, HL04239, the American Heart Association, the Department of Defense Myeloproliferative Disorders Research Program, a Stony Brook Targeted Research Award, and NIH Center grant MO1 10710–5 to the University Hospital General Clinical Research Center.

Prepublished online July 13, 2005 DOI: 10.1160/TH05-01-0037

platelet profiling using apheresis techniques (1, 6), prompting us to adapt this method for the molecular study of ET. We now demonstrate that human blood platelets express a previously described (testis-specific) enzyme (type 3 17β-hydroxysteroid dehydrogenase) known to catalyze the penultimate step in testosterone interconversion; furthermore, distinct expression patterns of genes encoding this enzyme are associated with the ET phenotype. These data identify establish a novel link between platelet function and hormone steroidogenic pathways that may be causally implicated in platelet-mediated thrombotic events, and/or platelet production.

Methods

Patient selection and characterization

Patients were enrolled from the larger pool of patients referred to the Division of Hematology for evaluation of thrombocytosis. All patients provided informed consent for an IRB (Institutional Review Board)-approved protocol completed in conjunction with the Stony Brook University Hospital General Clinical Research Center. Standard hematological criteria were followed for the diagnosis of essential thrombocythemia, reactive thrombocytosis, or other myeloproliferative disorders (7, 8). Both sex- and age-distribution paralleled prevalence figures for ET, with a M:F ratio of 1:2.3, and age at diagnosis ranging 23–78 years. Platelet counts at the time of blood isolation ranged from normal (reflecting treatment) to 1,724,000/μl; patient utilization of platelet-lowering drugs (i.e. hydroxyurea, analgrelide, or untreated) was recorded at the time of platelet isolation and purification (refer to Table 1 for detailed patient characteristics).

Platelet molecular studies

Platelets were obtained by apheresis or from peripheral blood (10 ml), and were isolated essentially as previously described, utilizing gel-filtration and CD45-coupled magnetic micro-beads for leukocyte immunodepletion (1). The final platelet-enriched product contained no more than 3–5 leukocytes per 1 X 10^5 platelets; peripheral blood leukocytes from three healthy donors were isolated as previously described (1). Pure cellular pellets were resuspended in 10 ml of Trizol reagent (Invitrogen, Carlsbad, CA), transferred into DEPC (diethylpyrocarbonate)-treated Corex tubes, and serially purified and precipitated using isopropanol (9). Platelet total mRNA quantification and integrity were established using an Agilent 2100 Bioanalyzer, and mean platelet RNA concentrations between the two groups were nearly identical: ET platelets contained ~0.8 ± 0.2 fg/platelet while normal platelet RNA concentrations were ~0.6 ± 0.3 fg/platelet. Quantitative reverse transcription (RT)-PCR was performed using fluorescence-based real-time PCR technology (TaqMan Real-Time PCR, Applied Biosystems, Foster City, CA). Oligonucleotide primer pairs were generated using Primer3 software (www-genome.wi.mit.edu), designed to generate ~200-bp PCR products. *HSD17B3*-specific primers were: forward (5'-3' AAATGTGATAACCAAGACTGC [bp 755–775]; reverse (5'-3' CTTGGTGTTGAGCTTCAGGTA [bp 956–936]; *HSD17B12*-specific primers were: forward (5'-3' TGAATACTTTTGGATGTTCCTGA [bp 496–519]); reverse (5'-3' AGTCTTGGTTGCAGAATAGATGGT, [bp 634–611]; *HSD17B11*-specific primers were: forward (5'-3' TGGATATAAAATGAAAGCG CAATA [bp 1067–1090]; reverse (5'-3': ATCAGCTTTTG GCTAAAGAACAAG [bp 1265–1242]; *F7*-specific primers were: forward (5'-3': TCCTGTTGTTGGTGAATGG [bp 734–753]; reverse (5'-3': GTACGTGCTGGGGATGATG [bp933–915]; β-actin-specific primers were as previously described (1). Purified platelet mRNA (4 μg) was used for first strand cDNA synthesis using random hexamers and SuperScript II reverse transcriptase (Invitrogen, Carlsbad, CA). For RT-PCR analysis, the RT reaction was equally divided among primer pairs and used in a 40-cycle PCR reaction for each target gene using three-step cycles of 94°C for 30 seconds, 55°C for 30 seconds, and 72°C for 1 min; mRNA levels were quantified by monitoring real-time fluorimetric intensity of SYBR green I. Relative mRNA abundance was determined from triplicate assays performed in parallel for each primer pair, calculated and standardized to β-actin as previously described (6, 10). For some patients, high molecular-weight genomic DNA was isolated from peripheral blood leukocytes for PCR-based amplification and sequencing of exons and exon-intron boundaries (11).

Gene expression profiles were completed using the HU133A GeneChip containing a 22, 283 probe set (Affymetrix, Santa Clara, CA). Total cellular RNA (5.8 μg) was used for cDNA syn-

Table 1: Patient characteristics.

ID	Age[1]	Sex	Diagnosis	Sample Source[2]	Platelet count[3] (X 10^9/L)	Treatment[4]	Genetic analysis[5]
ET1	31	F	ET	B,P	1,308	H	M,P
ET3	49	M	ET	B,P	565	H	M,P
ET4	33	M	ET	B,P	1,515	N	M,P
ET5	23	F	ET	B,P	1,566	N	M,P
ET6	37	F	ET	B,P	539	A	M,P
ET7	65	F	ET	B	991	A	P
ET9	64	F	ET	B	588	A	P
ET10	57	M	PV/ET6	B	945	H	P
ET11	71	F	ET	B	477	H	P
ET12	40	F	ET	B	456	N	P
ET14	65	F	ET	B	853	A	P
ET15	64	F	ET	B	511	H	P
ET18	78	M	ET	B	329 (940)	H	P
ET19	77	F	ET	B	488	H	P
ET20	69	M	ET	B	345 (1,063)	A	P
ET27	50	F	PV/ET6	B,P	1,724	H	M,P
ET28	29	F	RT7	B	402	N	P
ET29	66	F	ET	B	495	N	P
ET31	70	M	ET	B	347 (880)	A	P
ET32	50	F	ET	B	513	N	P
ET33	75	F	ET	B	641	H	P

[1]Age at diagnosis
[2]Sample source: B – peripheral blood; P – plateletpheresis
[3]Platelet count at time of sample collection (normal range 150 – 350); note that for patients with normal platelet counts at time of blood isolation, the platelet counts in parentheses are highest pre-treatment determinations
[4]Refers to treatment at the time of platelet isolation; A – Anagrelide; H- Hydroxyurea; N – Not treated 5M – Microarray; P – Quantitative RT-PCR
[6]Originally given the diagnosis of polycythemia rubra vera (PV)
[7]Secondary (post-splenectomy) thrombocytosis

thesis using SuperScript Choice system (Life Technologies, Rockville, MD) and an oligo (dT) primer containing the T7 polymerase recognition sequence, followed by cDNA purification using phenol/chloroform extraction and ethanol precipitation (1). *In vitro* transcription was completed in the presence of biotinylated ribonucleotides using a BioArray HighYield RNA Transcript Labeling Kit (Enzo Diagnostics, Farmingdale, NY), and after metal-induced fragmentation, 10 μg of the biotinylated cRNA was hybridized to the GeneChip array for 16 hours at 45°C. After washing, the cRNA was detected with streptavidin-phycoerythrin (Molecular Probes, Eugene, OR) and analysis completed using a Hewlett-Packard Gene Array Scanner. The fluorescence intensity of each probe was quantified using Affymetrix GeneChip software (MAS version 5.0), calculated as an average difference for each gene set obtained from 16 to 20 paired (perfectly matched and single nucleotide-mismatched) 25-bp oligonucleotides. The software is designed to exclude "positive calls" in the presence of high average differences with associated high mismatch intensities.

Bioinformatic and statistical analyses

Microarray data were analyzed and visualized using GeneSpring (version 7.0) software (Silicon Genetics, a subsidiary of Agilent Technologies, Palo Alto, CA). Megakaryocyte expression data were obtained from the National Center for Biotechnology (NCBI) Gene Expression Omnibus (GEO) database (http://www.ncbi.nlm.nih.gov/projects/geo/; platform GPL96), using profiles obtained from thrombopoietin-stimulated normal and ET bone marrow $CD34^+$ cells differentiated in liquid culture for 14 to 16 days (12). $CD34^+$-derived megakaryocytes (Mks) were purified using a megakaryocyte-specific anti-CD41a monoclonal antibody, followed by microarray analysis using the identical HU133A GeneChip. Data were normalized by dividing each measurement by the 50th percentile of all measurements in that sample, and each gene was divided by the median of its measurements in all samples. Normalized median ratios of individual genes were filtered for presence across arrays, and selected for expression levels as detailed. Prior to unsupervised hierarchical clustering of the uncentered Pearson correlation similarity matrix, the platelet microarray data were filtered for gene expression across phenotypic cohorts, defined as those genes present or marginal in a minimum of 80% of platelet samples (yielding 2,906 transcripts). A subset of genes was culled from the 2,906-gene list to specifically delineate those transcripts uniquely expressed in platelets; this platelet-restricted subset (N=126) was delineated by removing genes expressed in 3/3 leukocyte microarrays. A non-parametric analysis of variance test (ANOVA) was performed to identify differentially expressed genes using the Benjamini and Hochberg method to lower the false discovery rate ($p < 0.01$). All statistical analyses were completed using SPSS (Statistical Package for Social Sciences, version 11.5) software.

Functional 17β–HSD studies

Functional studies for platelet 17β–HSD3 activity were completed using gel-filtered platelets (GFP). Briefly, 1.5 X 10^8 platelets were solubilized and freeze-thawed in HSD buffer containing {20 mM Tris-HCl (pH 8.0), 5 mM $MgCl_2$, 50 mM NaCl, 5% Glycerol, 10 mM DTT, 1.5 mM NAD (D-5755, Sigma Co., St. Louis, MO)}, and the reaction started by addition of 1 μl (12.7 pmol) of [1,2,6,7-^{3}H(N)]-testosterone (specific activity 78.5 Ci/mM) (Dupont/NEN) to equivalent protein aliquots. A 30– or 90-minute reaction was allowed to proceed at 25^0 C, and quenched at –20°C in the presence of cold testosterone (1 mg/ml) and androstenedione (1 mg/ml). Steroid extraction was performed twice using 150 μl of ethyl acetate, both fractions were combined and air-dried under a vacuum. Androstenedione and testosterone were separated by thin layer chromatography (Uniplate Silica Gel GF plates (250 μm), Analtech Inc, Newark, DE) using 4:1 (v/v) chloroform:ethyl acetate. Discrete steroid fractions were visualized and extracted from silica gels using ethanol, and quantified by liquid scintillography. Purified hydroxysteroid dehydrogenase from *Pseudomonas testosteroni* (Sigma Co., St. Louis, MO) diluted to 10 mg/ml in HSD buffer served as control for some experiments. Mouse testis extract prepared from a C57/Bl6 mouse served as standard for platelet 17β-HSD3 quantification, and was prepared by homogenization in 3 ml of HSD buffer. After centrifugation at 5000 g for 5 minutes, supernatants were gel-filtered in HSD-equilibrated Centrisep spin columns prior to use; protein quantification of all samples was determined by optical density as previously described (13).

Results

The genetic profiles of highly-purified apheresis platelets isolated from 6 ET patients (4 females, 2 males) and 5 normal, healthy controls demonstrated distinctly different molecular signatures (Fig. 1). ET platelets collectively demonstrated higher numbers of expressed transcripts compared to normal controls, but considerably less than the transcript numbers generally found in nucleated cells (1). Of the genes classified as marginal or present in a minimum of 4 microarrays, ET patient samples expressed an average of 3,562 transcripts compared to 1,668 for normal controls (compared with ~10,500 transcripts identified in all three leukocyte microarrays). More stringent analyses (i.e. marginal or present in all of the arrays within a single group) extended these differences, with 1,840 transcripts expressed in ET platelets *versus* 1,086 transcripts expressed in platelets from healthy controls ($p < 0.03$). Thus, while bone marrow megakaryocyte expansion is known to accompany the thrombocythemic phenotype (2), this cellular proliferation is also associated with a nearly 45% increase in overall gene transcription.

An unsupervised, hierarchical clustering algorithm was used to group normal and ET platelet genes on the basis of similarities of gene expression (Fig. 1a,b). This direct comparison against a genetically normal platelet pool highlighted genes that consistently distinguish normal from diseased platelets (14). Normal or ET platelets have genetic profiles distinct from leukocytes, and notable differences between normal and ET platelets are clearly evident. To gain further insight into the specific changes associated with the ET phenotype – and to identify genes that could discriminate between morphologically indistinguishable ET and normal platelets – several methods of computational analyses were applied. Initial one-way ANOVA identified 170 genes that were differentially expressed, the majority of which (141) were up-regulated in ET platelets; only 29 genes were down-regulated

in ET compared to normal platelets (*see below*). Functional cluster analysis of this limited set of differentially-expressed genes (Fig. 2) demonstrated that genes involved in adhesion and catalytic activity represented the largest subgroups, although a sizable number of genes (35%) remained unclassified.

Because of inherent difficulties in analysis of microarray data sets (15), we re-analyzed the data, now computing *t*-statistics of ET *versus* normal platelets for each gene; the *t*-values were ranked by absolute magnitude (thereby incorporating inter-sample variability in expression ratios); they were then ranked by

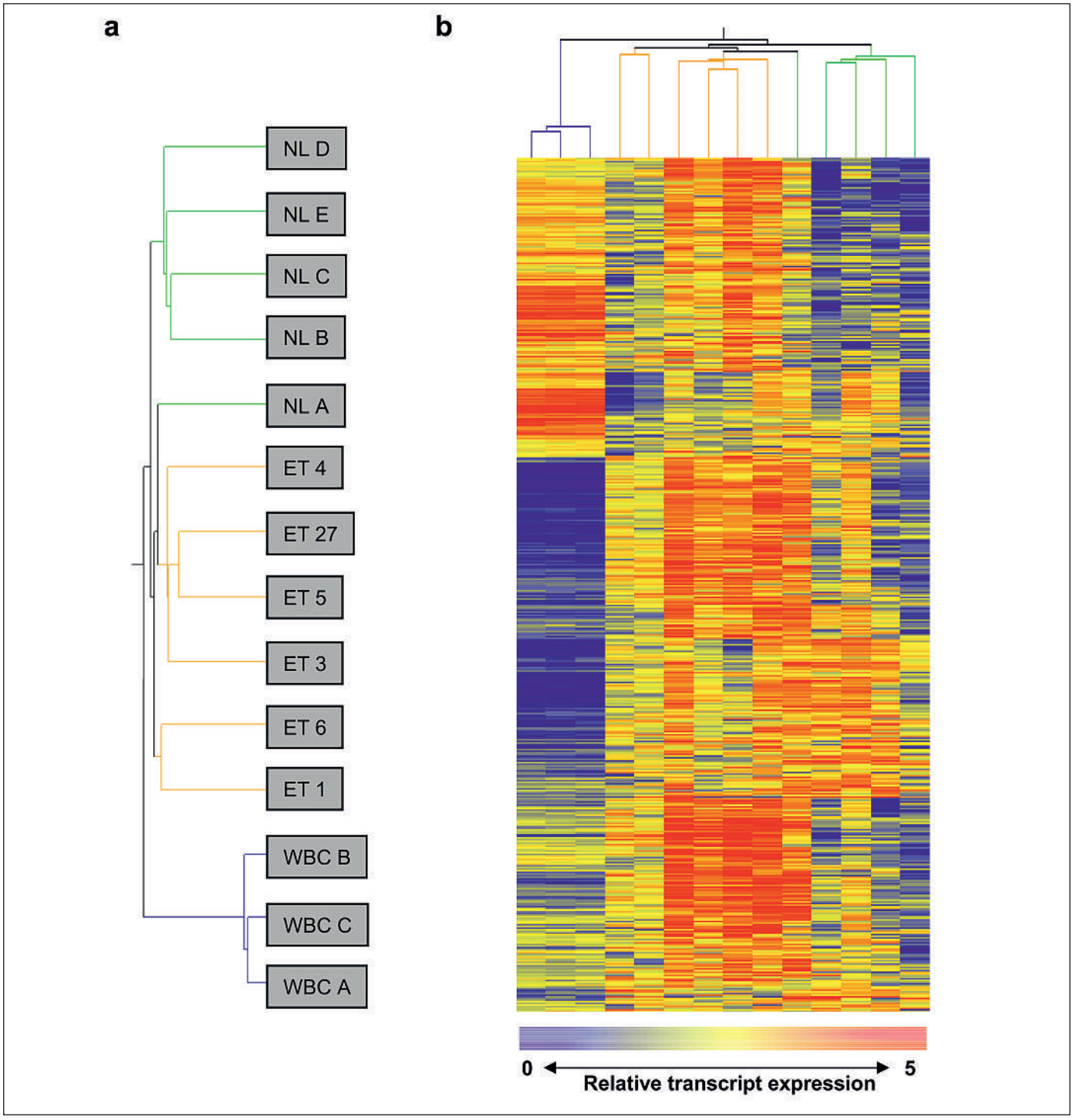


Figure 1: Molecular signature of normal and essential thrombocythemic platelets. Gene expression profiles from 14 experimental samples (eleven apheresis donors [5 normal, NL A-E; 6 patients with essential thrombocythemia (ET)], or 3 normal leukocyte [WBC A-C] donors) are displayed. a. Relationships between the experimental samples are displayed as dendrograms, in which the pattern and length of the branches depict sample cohort relatedness among the experimental groups. b. Unsupervised hierarchical clustering using the 2,906-gene set demonstrates the distinct variation in gene expression pattern among defined cohorts; each row represents a single gene.

the magnitude of the test statistic numerator, a measure of the biological difference in expression ratios. By applying a rigid 5-fold difference in pair-wise expression as the cut-off, we identified 163 genes that were up-regulated in ET, but only a small number (5) of down-regulated genes (Fig. 3a). To further pare this list, the analyses were repeated using the database of genes whose expression was restricted to platelets. This subset of platelet-restricted genes was delineated by excluding genes from the 2,906-gene list that were expressed in all 3 leukocyte arrays, leaving only 13 genes that were differentially-expressed (12 up-regulated, 1 down-regulated in ET)(Fig. 3b). Of the small subset of platelet-restricted, differentially expressed genes identified by this analysis, 7 were also in the top 40 list identified by the one-way ANOVA, establishing an independent layer of validation to these findings (Fig. 3c); while only 17% of the 170 differentially-expressed genes were under-expressed in ET platelets by one-way ANOVA, down-regulated genes were over-represented in the platelet-restricted gene list, accounting for 36% (9/25) of the total. Of the top 20 over-expressed genes, 3 encoded proteases or protease inhibitors (*HPSE*, *MMP1*, *SERPINI1*), a class of proteins well-associated in tumor invasiveness and cancer metastases (16). Interestingly, matrix metalloproteinases have recently been shown to mediate megakaryocyte transendothelial migration and proplatelet formation, although this effect appeared restricted to MMP-9 to the exclusion of MMP-2 (17); while the latter report did not specifically study MMP-1, our data would implicate MMP-1 as having a specific or ancillary role in the basement degradative process known to accompany proplatelet formation.

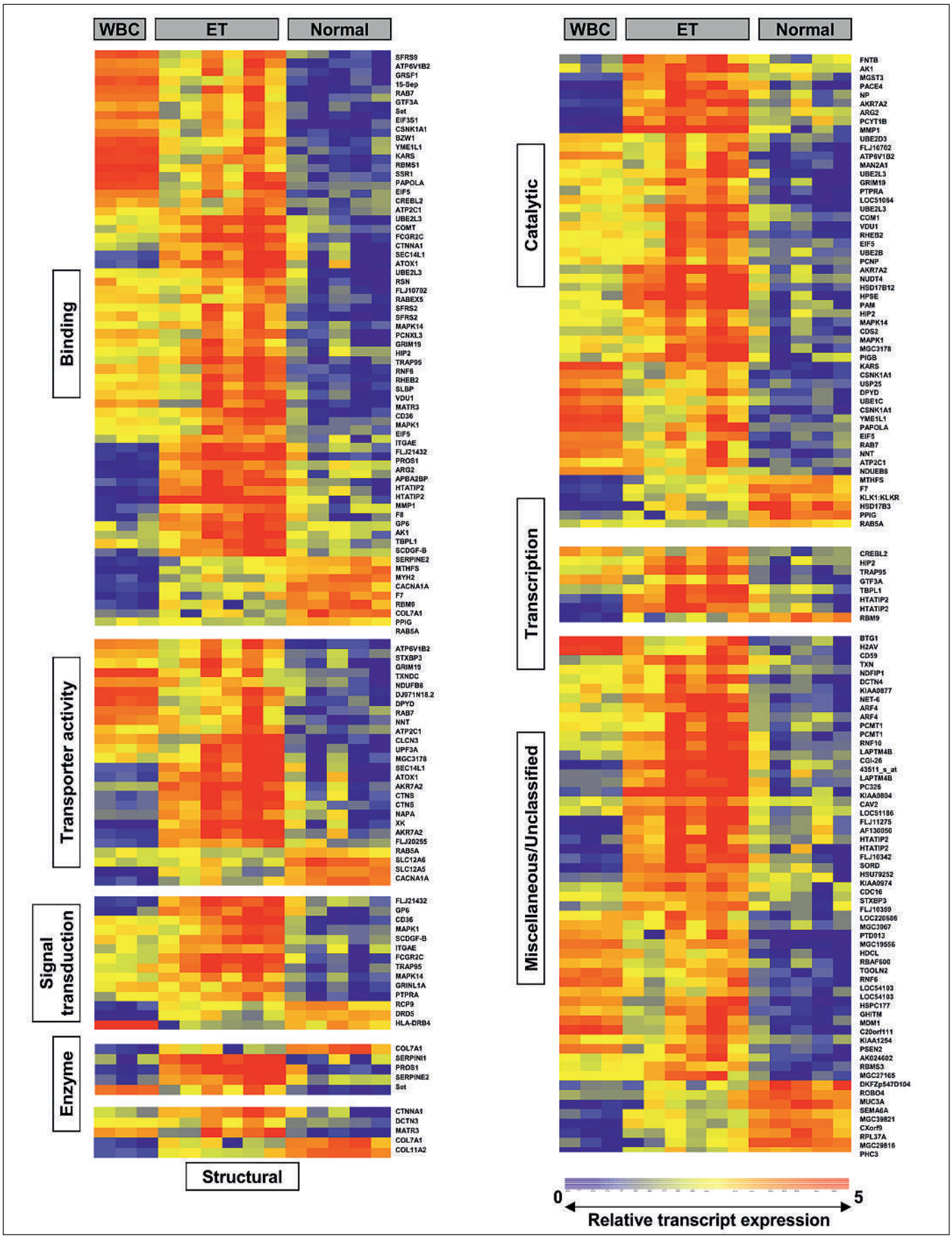
A single platelet-restricted gene *HSD17B3* (encoding the type 3 17β-hyrdroxysteroid dehydrogenase [17β-HSD3]) was expressed in all normal platelet arrays, and uniquely under-expressed in ET compared to normal platelets (Fig. 3b). The large family of steroid dehydrogenases retain oxidoreductase activity in a wide range of biological processes, although the 17β-HSDs uniquely function in the formation and inactivation of all active androgens and estrogens, with substrate-product interconversion regulated by the oxidative state of the NADP/NAD(P)H cofactors (18). The HSD type 3 enzyme is generally regarded as testes-specific, although rare SAGE (19) tags have been identified in CGAP tissues including brain (7 X 10^{-6}), skeletal muscle (9.3 X 10^{-6}), and prostate (1.6 X 10^{-5}) (20). The 17β-HSD3 enzyme specifically mediates the catalytic interconversion involving 4-androstenedione and testosterone, and molecular defects of the *HSD17B3* gene are causally implicated in male pseudohermaphroditism (21). While steroidogenic pathways are incompletely characterized in platelets, previous data have demonstrated that megakaryocytes (Mk) express the glucocorticoid receptor, and that both Mk and platelets selectively express estrogen receptor β (ERβ) and androgen receptor mRNA and protein, to the exclusion of ER α or progesterone receptor (22). Furthermore, Mk express functional 3β-HSD that is known to catalyze the essential step in the transformation of 5-pregnen-3β-ol and 5-androsten-3β-ol steroids into the corresponding Δ^4–3-keto-steroids, i.e. progesterone as well as the precursors of all androgens, estrogens, glucocorticoids and mineralocorticoids. Indeed, Mk-derived estradiol triggers megakaryocyte proplatelet formation *in vitro*, a process that is blocked through inhibition of 3β-HSD activity (23).

Figure 2: Functional clusters of differentially expressed genes from 14 experimental samples (see legend to Fig. 1). Gene Ontology classifications are delineated for the differentially-expressed gene set (N=170) as identified by one-way ANOVA ($p<0.01$).

To date, genes encoding twelve types of 17β-HSD enzymes have been described, although the type 6 and 9 genes have been only characterized in rodents (24). Oligonucleotide probes specific for all ten *HSD17B* genes are represented on the Affymetrix HU133a gene chip; probes for type 6 and 9 are not represented. Examination of the microarray data demonstrated that platelet *HSD17B* transcript expression was limited to three isoforms: *HSD17B3*, *HSD17B11*, and *HSD17B12*. While the *HSD17B11* mean, normalized signals between the normal and diseased cohorts were low-level and not statistically different, there was a striking change in the pattern of *HSD17B3* and *HSD17B12* expression between ET and normal platelets (Fig. 4 a-c). Absence of *HSD17B3* transcript expression was evident in all 6 ET patients, changes that occurred concomitantly with elevated transcript levels of *HSD17B12* in the same patient subgroup. In contrast, expression of *HSD17B3* in normal platelets was accompanied by negligible to low-level *HSD17B12* expression. To compare these platelet profiles with those from ET and normal megakaryocytes, we downloaded Mk profiles from the NCBI GEO database for bioinformatic analyses using these published data sets (12). ET Mks demonstrated the identical patterns as those found in ET platelets, i.e. absence of *HSD17B3* expression with enhanced expression of the *HSD17B12* gene. Somewhat unexpectedly, these same patterns were also evident in normal Mks; while these comparisons are of potential relevance, they are nonetheless limited by the exogenous cytokine supplementation required for *ex vivo* Mk differentiation.

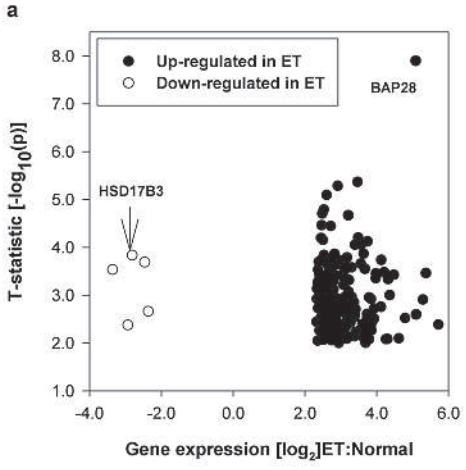
To validate and extend these findings, we developed a quantitative RT-PCR (qRT-PCR) assay, and applied this assay to the original ET cohort and an expanded cohort of normal controls, specifically collected to exclude potential gender-bias in *HSD17B* gene expression. These results confirmed and paralleled those found by microarray, demonstrating ~4.5-fold greater *HSD17B3* transcript levels in normal platelets (compared to ET, $p \leq 0.001$) and concomitant ~27-fold greater *HSD17B12* transcript expression in ET platelets (compared to normal, $p \leq 0.03$); these reciprocal changes amount to an aggregate ~2-log change in intracellular *HSD17B3*:*HSD17B12* transcript levels between normal and ET platelets (Fig. 4d). Since the qRT-PCR data established an absolute decrease in *HSD17B3* transcript level in ET platelets, preliminary genomic analyses of *HSD17B3* were completed in 4 of the 6 ET patients. The 11 exons and intron-exon boundaries were amplified and sequenced, resulting in identification of a single heterozygous A insertion [not involving the splice junction site (25)] in the first intron of one patient (ET1). Thus, there was no evidence that a small deletion or missense mutation affecting *HSD17B3* transcript stability was causally implicated in reduced *HSD17B3* transcript expression (data not shown).

The distinct patterns of *HSD17B* expression identified by microarray and confirmed by qRT-PCR were then extended to a


WBC
ET
Normal
Binding
SFRS9
ATP6V1B2
GRSF1
15-Sep
RAB7
GTF3A
Set
EIF3S1
CSNK1A1
BZW1
YME1L1
KARS
RBMS1
SSR1
PAPOLA
EIF5
CREBL2
ATP2C1
UBE2L3
COMT
FCGR2C
CTNNA1
SEC14L1
ATOX1
UBE2L3
RSN
FLJ10702
RABEX5
SFRS2
SFRS2
MAPK14
PCNXL3
GRIM19
HIP2
TRAP95
RNF6
RHEB2
SLBP
VDU1
MATR3
CD36
MAPK1
EIF5
ITGAE
FLJ21432
PROS1
ARG2
APBA2BP
HTATIP2
MMP1
F8
GP6
AK1
TBPL1
SCDGF-B
SERPINE2
MTHFS
MYH2
CACNA1A
F7
RBM9
COL7A1
PPIG
RAB5A
Transporter activity
ATP6V1B2
STXBP3
GRIM19
TXNDC
NDUFB8
DJ971N18.2
DPYD
RAB7
NNT
ATP2C1
CLCN3
UPF3A
MGC3178
SEC14L1
ATOX1
AKR7A2
CTNS
CTNS
NAPA
XK
AKR7A2
FLJ20255
RAB5A
SLC12A6
SLC12A5
CACNA1A
Signal transduction
FLJ21432
GP6
CD36
MAPK1
SCDGF-B
ITGAE
FCGR2C
TRAP95
MAPK14
GRINL1A
PTPRA
RCP9
DRD5
HLA-DRB4
Enzyme
COL7A1
SERPINI1
PROS1
SERPINE2
Set
Structural
CTNNA1
DCTN3
MATR3
COL7A1
COL11A2
WBC
ET
Normal
Catalytic
FNTB
AK1
MGST3
PACE4
NP
AKR7A2
ARG2
PCYT1B
MMP1
UBE2D3
FLJ10702
ATP6V1B2
MAN2A1
UBE2L3
GRIM19
PTPRA
LOC51084
UBE2L3
COM1
VDU1
RHEB2
EIF5
UBE2B
PCNP
AKR7A2
NUDT4
HSD17B12
HPSE
PAM
HIP2
MAPK14
CDS2
MAPK1
MGC3178
PIGB
KARS
CSNK1A1
USP25
DPYD
UBE1C
CSNK1A1
YME1L1
PAPOLA
EIF5
RAB7
NNT
ATP2C1
NDUEB8
MTHFS
F7
KLK1:KLKR
HSD17B3
PPIG
RAB5A
Transcription
CREBL2
HIP2
TRAP95
GTF3A
TBPL1
HTATIP2
HTATIP2
RBM9
Miscellaneous/Unclassified
BTG1
H2AV
CD59
TXN
NDFIP1
DCTN4
KIAA0877
NET-6
ARF4
ARF4
PCMT1
PCMT1
RNF10
LAPTM4B
CGI-26
43511_s_at
LAPTM4B
PC326
KIAA0804
CAV2
LOC51186
FLJ11275
AF130050
HTATIP2
HTATIP2
FLJ10342
SORD
HSU79252
KIAA0974
CDC16
STXBP3
FLJ10359
LOC220686
MGC3067
PTD013
MGC19556
HDCL
RBAF600
TGOLN2
RNF6
LOC54103
LOC54103
HSPC177
GHITM
MDM1
C20orf111
KIAA1254
PSEN2
AK024602
RBMS3
MGC27165
DKFZp547D104
ROBO4
MUC3A
SEMA6A
MGC39821
CXorf9
RPL37A
MGC29816
PHC3
0
5
Relative transcript expression

larger cohort of 20 ET patients (6 original ET patients and 14 newly-studied individuals), now uniformly analyzed using peripheral blood as the starting source for platelet analysis (bypassing the need for cumbersome apheresis technique used in the original cohorts). The qRT-PCR results were entirely concordant for all individuals studied, demonstrating that *17BHSD12:17BHSD3* transcript ratios reliably predicted the ET phenotype in all patients studied to date ($p < 0.0001$) (Fig. 5). Furthermore, these differential patterns of *HSD17B* expression appeared unrelated to the development of thrombocytosis *per se*, but rather, were restricted to the ET phenotype. This observation is based on the results of four individuals: ET28 who has secondary thrombocytosis and ratios predictive of the normal phenotype; and ET18, ET20, and ET31 with aggressively-treated ET and normal platelet counts who maintain *17BHSD12:17BHSD3* transcript ratios predictive of ET. The causes of thrombocytosis are varied, and larger cohorts of patients with etiologically diverse causes for thrombocytosis will need to be studied for confirmation, and for potential discriminatory value. Nonetheless, these initial data suggest that the intracellular signals regulating MK/platelet *17BHSD* expression may be associated with the pathophysiological mechanism(s) of the ET phenotype.

To confirm that platelets retained functional 17β–HSD3 activity (and to compare this activity between normal and ET platelets), we quantified the oxidative conversion of testosterone to 4-androstenedione. Entirely consistent with the gene expression data, normal platelets retained 17β–HSD3 activity, providing (to our knowledge) the first evidence that non-testicular sources retain functional capacity in the penultimate step of androgen biosynthesis; furthermore, the platelet-derived 17β-HSD3 activity was not inconsequential, providing nearly 10% of the capacity found in testis (Fig. 6). Finally, in the initial cohort of patients studied, ET platelets demonstrated total 17β–HSD3 activity that was not statistically different from that found in normal platelets.



c

Gene Symbol	Gene Description	(ET:Normal)
HPSE	Heparanase	52.657
LAPTM4B	**Lysosomal associated protein transmembrane 4 beta**	41.344
FLJ10359	Protein BAP28	33.924
PROS1	Protein S (alpha)	22.212
SEC14L1	SEC14-like 1 (*S. cerevisiae*)	20.559
SERPINI1	Protease inhibitor 12 (neuroserpin)	19.707
FCGR2C	Fc fragment of IgG, low affinity IIb, receptor for (CD32)	19.104
NDFIP1	Nedd4 family interacting protein 1	19.007
CLCN3	Chloride channel 3	18.609
ATOX1	**ATX1 antioxidant protein 1 homolog (*yeast*)**	18.456
EIF3S1	Eukaryotic translation initiation factor 3, subunit 1 alpha	17.322
UPF3A	UPF3 regulator of transcripts homolog A (*S. cerevisiae*)	17.213
KIAA1254	Phosphatidylinositol glycan, class B	16.168
XK	Kell blood group precursor (McLeod phenotype)	15.777
MMP1	**Matrix metalloproteinase-1**	15.393
KIAA0877	KIAA0877 protein	15.184
HSD17B12	Hydroxysteroid (17-beta) dehydrogenase 12	14.522
PCMT1	Protein-L-isoaspartate (D-aspartate) O-methyltransferase	14.449
UBE2L3	Ubiquitin-conjugating enzyme E2L 3	14.039
TRAP95	Thyroid hormone receptor-associated protein, 95-kD subunit	13.416
HSD17B3	**Hydroxysteroid (17-beta) dehydrogenase 3**	0.142
MUC3A	Mucin 3A, intestinal	0.225
PSEN2	Presenilin 2 (Alzheimer disease 4)	0.229
RBMS3	**RNA binding motif, single stranded interacting protein**	0.235
PPIG	Peptidyl-prolyl isomerase G (cyclophilin G)	0.268
COL7A1	Collagen, type VII, alpha 1	0.297
RPL37A	Ribosomal protein L37a	0.301
SLC12A6	Solute carrier family 12 (K/Cl transporters), member 6	0.341
CXorf9	Human DNA sequence on Xq25-26.1	0.353
CACNA1A	Alpha1A-voltage-dependent calcium channel	0.363
RBM9	**RNA binding motif protein 9**	0.374
COL11A2	Collagen, type XI, alpha 2	0.376
(AK024602)	Transcribed locus, similar to gasdermin domain	0.380
SLC12A5	**Solute carrier family 12, (K-Cl transporter) member 5**	0.381
MGC29816	Hypothetical protein MGC29816	0.409
ROBO4	Roundabout homolog 4, magic roundabout (*Drosophila*)	0.414
DKFZp547D104	Acyl-CoA synthetase long-chain family member 6	0.415
HLA-DRB4	Major histocompatibility complex, class II, DR beta 4	0.418
F7	Coagulation factor VII	0.472
PHC3	Polyhomeotic like 3 (*Drosophila*)	0.501

Figure 3: Identification of most significant differentially-expressed ET genes. Gene selection was calculated for each gene by applying a 5-fold cut-off and a computed t-statistic, using the aggregate (N=2906) gene list (a), or the list of platelet-restricted (N=126) genes (b). c. The list of 40 genes (20 over-expressed, 20 under-expressed) demonstrating greatest differential expression by one-way ANOVA is presented, ranked by the ratio of the mean group normalized signals (ET:Normal); shaded column delineates genes with highest expression in ET whereas unshaded column refers to genes with lowest expression. Genes are delineated by gene symbol, or if unassigned, are specified by GenBank ID in parentheses. Genes in bold are platelet-restricted; 2/12 genes in Panel b (*ATOX1*, *LAPTM4B*) are not depicted because they are present in Panel c.

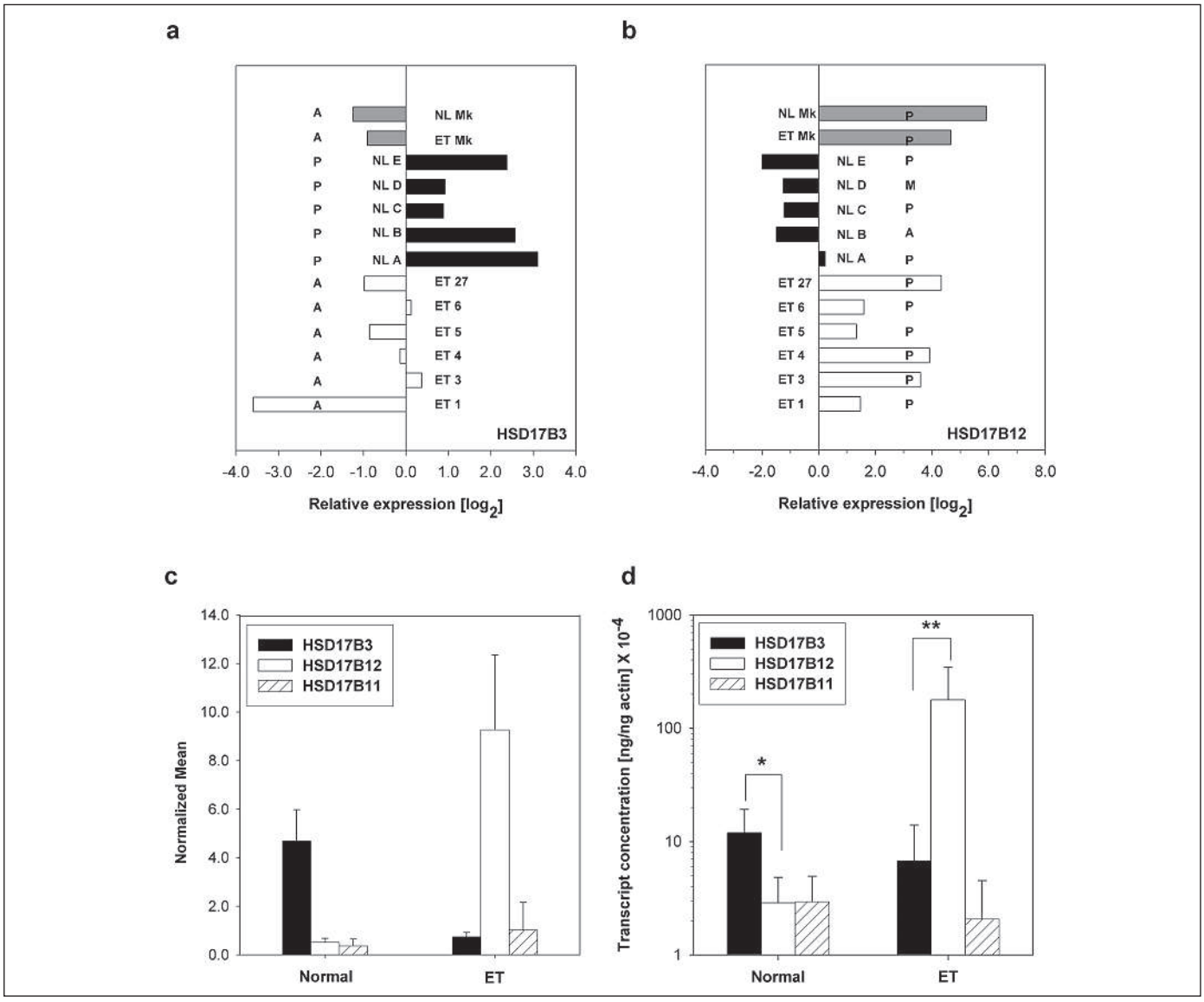


Figure 4: Transcript analysis of platelet-expressed *HSD17Bs*. Normalized microarray values for individual patients (Panels a, b) or the normalized aggregate means by group (c) are shown. In Panels a (*HSD17B3*) and b (*HSD17B12*), expression levels are $\log_2$-transformed, such that negative numbers reflect decreased expression compared to the normalized mean of 14 chips (6 ET, 5 normals, 3 WBC); for individual patients, delineation of transcript expression using Affymetrix MAS 5.0 software is specified (P – *present*; M – *marginal*; A- *absent*). The megakaryocyte (Mk) data in (a) and (b) were downloaded from the GEO platform GPL96 database, and analyzed using GeneSpring software. d. Quantitative RT-PCR was completed on the original cohort of 6 ET patients and a new cohort of 10 normal controls (5 males, 5 females) using *HSD17B3*-, *HSD17B11*-, *HSD17B12*-, or *F7*-specific oligonucleotide primers (platelet coagulation factor VII (*F7*) is primarily endocytosed from plasma with negligible to no platelet mRNA expression (34), thereby establishing the lower limit of assay sensitivity [$1.1 \times 10^{-5} \pm 2.6 \times 10^{-7}$ ng/ng actin, *not shown*]). *p ≤0.03; **p ≤0.001.

Thus, the high level expression of *HSD17B12* transcript in ET platelets is unassociated with overall enzymatic capabilities in androgen biosynthesis. To date, the potential role of 17β–HSD12 in human steroidogenesis remains unknown, although our data would suggest that its substrate specificity is distinct from that of 17β–HSD3. Phylogenetic analysis of the recently identified *HSD17B3* and *HSD17B12* orthologs from the zebrafish *Danio rerio* (26) suggest an evolutionarily conserved and potentially important physiological role for *HSD17B12*, although its precise function remains speculative and offers little insight to the function of its human ortholog.

Discussion

Our data provide the first evidence that distinct subtypes of the steroidogenic 17β–HSDs are functionally expressed in human blood platelets, and that the reciprocal expression patterns of *HSD17B3* and *HSD17B12* are distinctly associated with a rare platelet disorder manifest by quantitative and qualitative platelet defects. In all ET patients studied to date, a simple and reliable assay predicts the ET phenotype, providing a potentially diagnostic molecular marker for this disease; to date, we have insufficient evidence that this assay can distinguish ET from reactive

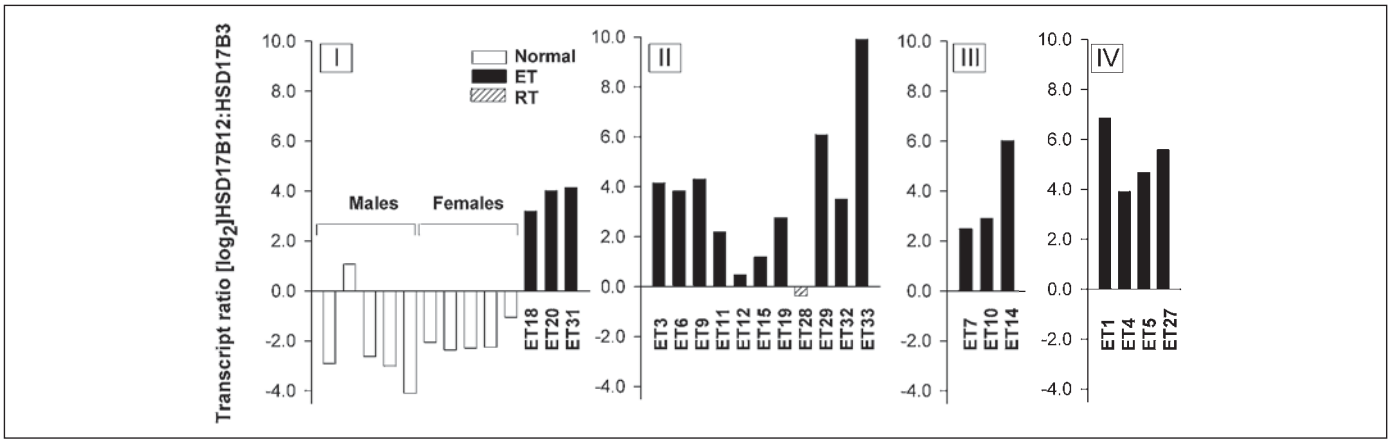


Figure 5: Genetic assays of normal and ET platelets. a. qRT-PCR was completed on platelets isolated by routine phlebotomy (10 ml) from ET or normal controls, using *HSD17B*-specific primers as outlined in Fig. 4d. The (log_2) ratio of *HSD17B12:HSD17B3* transcript expression is depicted for individual patients, categorized by platelet count (10^9/L) at time of platelet isolation; Category I: 0 – 350; Category II: 351–675; Category III: 676–999; Category IV: >1,000 (refer to Table 1 for detailed patient characteristics). Note that three Category I ET patient had normalized platelet counts with treatment at the time of analysis, and retained an *HSD17B12:HSD17B3* ratio predictive of the ET phenotype; the single patient with secondary thrombocytosis (RT) studied to date had a normal ratio.

thrombosis. Given the broad etiologies for thrombocytosis, a more comprehensive study is currently being designed for large cohort analysis and applicability.

What is the biological relevance of 17β–HSD expression in platelets, and is this causally associated with the molecular defect regulating megakaryocytopoiesis and/or proplatelet formation? Platelet production is complex, but is clearly affected by endogenous or exogenous sex hormones. Androgens improve platelet counts in patients with distinct types of thrombocytopenia (27), and diminished thrombopoiesis is evident in castrated mice, with correction upon testosterone supplementation (28). Furthermore, murine Mk retain the machinery for estradiol synthesis *in vitro*, and the secreted hormone's ability to bind estrogen receptor β triggers proplatelet formation in an autocrine manner (23). Although we did not characterize *HSD17B* expression in Mk, diminished platelet *HSD17B3* transcripts (with exaggerated *HSD17B12* expression) presumably reflects comparable changes in Mk. While *HSD17B3* transcript is down-

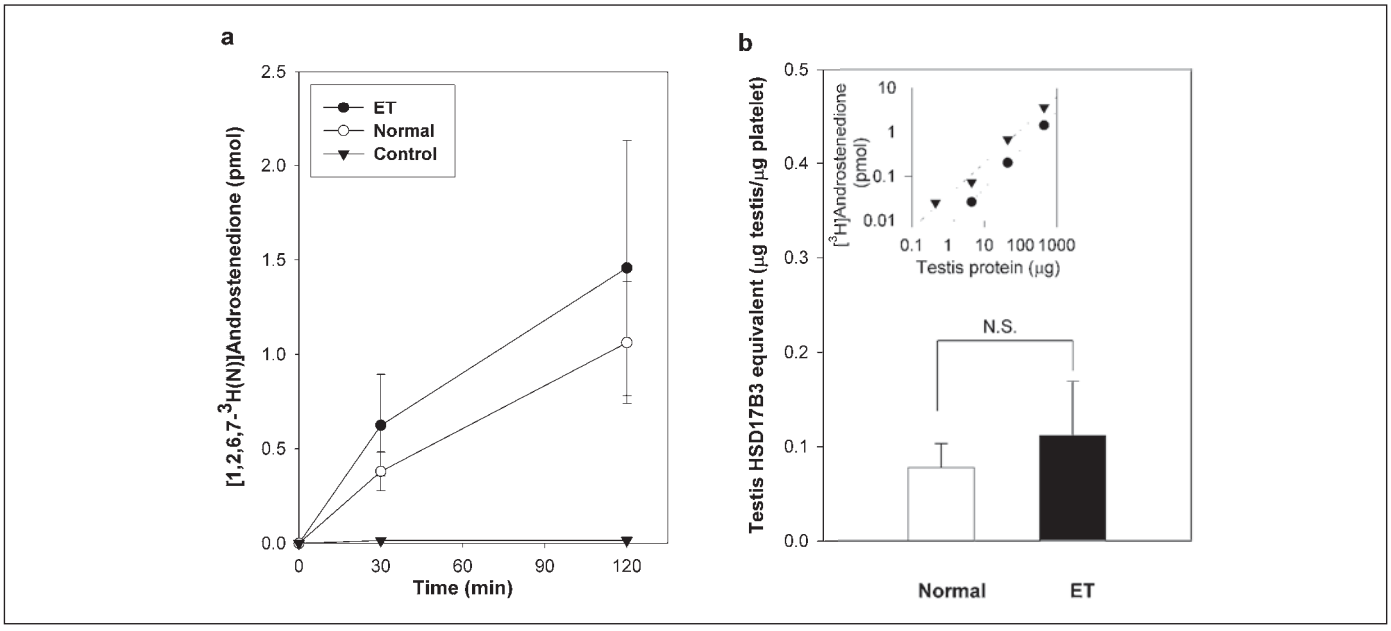


Figure 6: Functional 17β-HSD assays in platelets. Lysates of platelets isolated from 3 normal or 3 ET patients were used for determination of functional 17β-HSD3 activity at distinct time points, and expressed as standardized mean ± SEM (control is HSD buffer alone). In Panel b, curves generated for 30-minute (●) or 120-minute (▲) 17-βHSD3 assays (inset) were used as standards for comparative determination of platelet 17β-HSD3 assays performed in parallel. Data are the mean ± SEM from 6 determinations; note that 0.1 testis equivalent units correspond to 10% of the activity found in protein equivalent of mouse testes.

regulated in ET platelets, the changes are not as pronounced as those of *HSD17B12* where transcript expression is ~25-fold higher than in normal platelets. Despite this exaggerated increase, our data demonstrate that *HSD17B12* induction is unassociated with enzymatic changes in total platelet 17β–HSD3 activity. Thus, the function and substrate specificity of human 17β-HSD12 appears distinct from that of 17β–HSD3, and elucidation of its substrate(s) may provide further insights into the roles of 17β–HSDs in platelet and/or Mk functions.

Human platelets express both androgen and estrogen β receptors, and platelet function is known to be modulated by gender differences (29), the menstrual cycle (30), and exogenous testosterone (31). Similarly, hormonal replacement in postmenopausal females, or oral contraceptive use in menstruating females are known to predispose to thrombotic diseases (32, 33), although the mechanism(s) of this effect remain unclear. These effects could be mediated at either genomic (Mk transcriptional activity regulated by hormonal/receptor binding) or nongenomic (non-transcriptional) levels, although an interplay at both levels is equally plausible. The evidence that human blood platelets retain 17β-HSD3 activity, express distinct subtypes of 17β-HSD enzymes, and demonstrate altered *HSD17B* expression patterns in a disorder known to be association with thrombohemorrhagic risk, provides novel insights into the interplay between sex hormones, platelet function, and vascular diseases, both cerebro- and cardiovascular.

Acknowledgements

We would like to thank Dr. Eli Hatchwell (Cold Spring Harbor Laboratory) and Varsha Sitaraman for assistance with the genomic sequence analysis, Dr. Linda Tseng (Obstetrics/Gynecology, SUNY/Stony Brook) for assistance with the functional 17β-HSD assays, and special thanks to all the patients who graciously volunteered their time and effort for this study.

Accession numbers

All microarray data were submitted to the GEO database in MIAME-compliant form, reported under the accession number GPL1716.

References

1. Gnatenko DV, Dunn JJ, McCorkle SR et al. Transcript profiling of human platelets using microarray and serial analysis of gene expression. Blood 2003; 101: 2285–93.
2. Nimer SD. Essential thrombocythemia: Another "heterogeneous disease" better understood? Blood 1999; 93: 415–16.
3. El-Kassar N, Hetet G, Briere J et al. Clonality analysis of hematopoiesis in essential thrombocythemia: Advantages of studying T lymphocytes and platelets. Blood 1997; 89: 128.
4. Kondo T, Okabe M, Sanada M et al. Familial essential thrombocythemia associated with one-base deletion in the 5'-untranslated region of the thrombopoietin gene. Blood 1998; 92: 1091.
5. Schafer AI. Thrombocytosis. N Engl J Med 2004; 350: 1211–9.
6. Bahou WF, Gnatenko DV. Platelet transcriptome: the application of microarray analysis to platelets. Semin Thromb Hemost 2004; 30: 473–84.
7. Iland HJ, Laszlo J, Case DC, Jr. et al. Differentiation between essential thrombocythemia and polycythemia vera with marked thrombocytosis. Am J Hematol 1987; 25: 191–201.
8. Murphy S, Peterson P, Iland H et al. Experience of the Polycythemia Vera Study Group with essential thrombocythemia: a final report on diagnostic criteria, survival, and leukemic transition by treatment. Semin Hematol 1997; 34: 29–39.
9. Mirza H, Yatsula V, Bahou WF. The proteinase activated receptor-2 (PAR-2) mediates mitogenic responses in human vascular endothelial cells. Molecular characterization and evidence for functional coupling to the thrombin receptor. J Clin Invest 1996; 97: 1705–14.
10. Heid CA, Stevens J, Livak KJ et al. Real time quantitative PCR. Genome Res 1996; 6: 986–94.
11. Schmidt VA, Nierman WC, Maglott DR et al. The human proteinase-activated receptor-3 (PAR-3) gene. Identification within a PAR gene cluster and characterization in vascular endothelial cells and platelets. J Biol Chem 1998; 273: 15061–8.
12. Tenedini E, Fagioli ME, Vianelli N et al. Gene expression profiling of normal and malignant CD34-derived megakaryocytic cells. Blood 2004; 104: 3126–35.
13. Bahou WF, Scudder L, Rubenstein D et al. A shear-restricted pathway of platelet procoagulant activity is regulated by IQGAP1. J Biol Chem 2004; 279: 22571–7.
14. Dhanasekaran SM, Barrette TR, Ghosh D et al. Delineation of prognostic biomarkers in prostate cancer. Nature 2001; 412: 822–6.
15. Mehta T, Tanik M, Allison DB. Towards sound epistemological foundations of statistical methods for high-dimensional biology. Nat Genet 2004; 36: 943–7.
16. Bahou WF. Protease-activated receptors. Curr Top Dev Biol 2003; 54: 343–69.
17. Lane WJ, Dias S, Hattori K et al. Stromal-derived factor 1-induced megakaryocyte migration and platelet production is dependent on matrix metalloproteinases. Blood 2000; 96: 4152–9.
18. Labrie F, Luu-The V, Labrie C et al. Endocrine and intracrine sources of androgens in women: inhibition of breast cancer and other roles of androgens and their precursor dehydroepiandrosterone. Endocr Rev 2003; 24: 152–82.
19. Velculescu V, Zhang L, Vogelstein B et al. Serial analysis of gene expression. Science 1995; 270: 484–7.
20. Strausberg RL, Buetow KH, Emmert-Buck MR et al. The cancer genome anatomy project: building an annotated gene index. Trends Genet 2000; 16: 103–6.
21. Geissler WM, Davis DL, Wu L et al. Male pseudohermaphroditism caused by mutations of testicular 17 beta-hydroxysteroid dehydrogenase 3. Nat Genet 1994; 7: 34–9.
22. Khetawat G, Faraday N, Nealen ML et al. Human megakaryocytes and platelets contain the estrogen receptor beta and androgen receptor (AR): testosterone regulates AR expression. Blood 2000; 95: 2289–96.
23. Nagata Y, Yoshikawa J, Hashimoto A et al. Proplatelet formation of megakaryocytes is triggered by autocrine-synthesized estradiol. Genes Dev 2003; 17: 2864–9.
24. Mindnich R, Moller G, Adamski J. The role of 17 beta-hydroxysteroid dehydrogenases. Mol Cell Endocrinol 2004; 218: 7–20.
25. Mount S. A catologue of splice junction sequences. Nucl Acid Res 1982; 10: 459–72.
26. Mindnich R, Deluca D, Adamski J. Identification and characterization of 17 beta-hydroxysteroid dehydrogenases in the zebrafish, Danio rerio. Mol Cell Endocrinol 2004; 215: 19–30.
27. Wattel E, Cambier N, Caulier MT et al. Androgen therapy in myelodysplastic syndromes with thrombocytopenia: a report on 20 cases. Br J Haematol 1994; 87: 205–8.
28. Sullivan PS, Jackson CW, McDonald TP. Castration decreases thrombocytopoiesis and testosterone restores platelet production in castrated BALB/c mice: evidence that testosterone acts on a bipotential hematopoietic precursor cell. J Lab Clin Med 1995; 125: 326–33.
29. Johnson M, Ramey E, Ramwell PW. Sex and age differences in human platelet aggregation. Nature 1975; 253: 355–7.
30. Jones SB, Bylund DB, Rieser CA et al. alpha 2-Adrenergic receptor binding in human platelets: alterations during the menstrual cycle. Clin Pharmacol Ther 1983; 34: 90–6.
31. Pilo R, Aharony D, Raz A. Testosterone potentiation of ionophore and ADP induced platelet aggregation: relationship to arachidonic acid metabolism. Thromb Haemost 1981; 46: 538–42.
32. Castellsague J, Perez Gutthann S, Garcia Rodriguez LA. Recent epidemiological studies of the association between hormone replacement therapy and venous thromboembolism. A review. Drug Saf 1998; 18: 117–23.
33. Daly E, Vessey MP, Hawkins MM et al. Risk of venous thromboembolism in users of hormone replacement therapy. Lancet 1996; 348: 977–80.
34. Greenberg D, Miao CH, Ho WT et al. Liver-specific expression of the human factor VII gene. Proc Natl Acad Sci U S A 1995; 92: 12347–51.

Proteomic approaches to dissect platelet function: half the story

Dmitri V. Gnatenko, Peter L. Perrotta, and Wadie F. Bahou

Platelets play critical roles in diverse hemostatic and pathologic disorders and are broadly implicated in various biological processes that include inflammation, wound healing, and thrombosis. Recent progress in high-throughput mRNA and protein profiling techniques has advanced our understanding of the biological functions of platelets. Platelet proteomics has been adopted to decode the complex processes that underlie platelet function by identifying novel platelet-expressed proteins, dissecting mechanisms of signal or metabolic pathways, and analyzing functional changes of the platelet proteome in normal and pathologic states. The integration of transcriptomics and proteomics, coupled with progress in bioinformatics, provides novel tools for dissecting platelet biology. In this review, we focus on current advances in platelet proteomic studies, with emphasis on the importance of parallel transcriptomic studies to optimally dissect platelet function. Applications of these global profiling approaches to investigate platelet genetic diseases and platelet-related disorders are also addressed. (Blood. 2006;108:3983-3991)

Introduction

Human blood platelets play important roles in fundamental biological processes, including thrombosis, inflammation, wound repair, and stroke. Although they are anucleate and lack nuclear DNA, platelets retain small amounts of megakaryocyte-derived mRNA.[1,2] Platelets also contain rough endoplasmic reticulum and polyribosomes, thus retaining the capacity for protein biosynthesis from cytoplasmic mRNA.[3] Quiescent platelets display minimal translational activity, although platelet activation leads to the rapid translation of preexisting mRNA,[4] with the release or derivation of platelet-secreted proteins, cytokines, exosomes, and microparticles.

The traditional paradigm that platelet mRNA content is invariant and gradually declines with cell senescence was challenged when signal-dependent pre-mRNA splicing was identified in platelets.[5] Signal-dependent splicing provides a mechanism for altering the repertoire of translatable messages in response to cellular activation/stimulation. Furthermore, platelets have essential components of a functional spliceosome and selected unspliced pre-mRNAs. These spliceosomes retain a unique ability to splice pre-mRNA in the cytoplasm (as opposed to the typical nuclear location), a capability not described in any other mammalian cell. This discovery emphasizes that the molecular mechanisms of platelet function cannot be optimally dissected without accurate platelet transcript profiling.

Modern postgenomic, high-throughput approaches allow integrated studies of molecular components (at the RNA and the protein levels) involved in cell function. Platelets represent an attractive, simplified model for these studies because they lack nuclear DNA and because their genome consists of a small subset of megakaryocyte-derived mRNA transcripts. This complete pool of platelet RNAs is significantly smaller than the transcriptome of a nucleated cell.[6] The entire pool of platelet proteins constitutes the platelet proteome: the initially static, but functionally dynamic, protein interactions that occur with platelet activation. In this review, we focus on recent applications of proteomic and transcriptomic technologies to dissect platelet function in normal processes and pathologic disorders.

Platelet transcript profiling: transcriptomics

Modern approaches to transcript profiling

The development of global transcript profiling technologies, such as microarray and serial analysis of gene expression (SAGE),[7,8] provided novel methodologies for dissecting platelet function. Microarray analysis adapts artificially constructed grids of known DNA samples such that each element of the grid probes for a specific RNA sequence; these are then used to capture and quantify RNA transcripts.[9] Microarray platforms developed to date represent closed transcript profiling systems—that is, they detect only those transcripts that correspond to specific probes imprinted on the chip. Transcripts without corresponding probes are not detected. Recent technological advancements allow accurate whole genome transcript profiling and are capable of detecting alternatively spliced transcripts.[9]

SAGE represents an open transcript profiling system that can detect any transcript within the SAGE library. Classical SAGE[10] relies on the observation that short (less than 10 bp) sequences (tags) within 3′-mRNAs can stringently discriminate among the genes that constitute the human genome. Differentially expressed genes can be identified in a quantitative manner because the frequency of tag detection reflects the steady state mRNA level of the cellular transcriptome.[10,11] Genes expressed at low levels (less than 0.01% of total mRNA) can be identified by SAGE. Modified SAGE protocols have been devised to provide more definitive identification by using longer tags, identify low-abundant transcripts efficiently through subtractive SAGE techniques, and amplify small amounts of mRNA starting material.[12-14]

From the Department of Medicine and the Program in Genetics, State University of New York, Stony Brook, NY; and the Department of Pathology, West Virginia University, Morgantown, WV.

Submitted June 14, 2006; accepted August 2, 2006. Prepublished online as *Blood* First Edition Paper, August 22, 2006; DOI 10.1182/blood-2006-06-026518.

Analyzing the platelet transcriptome

Initial characterization of platelet-derived mRNA transcripts was achieved by constructing platelet-specific cDNA libraries[15] and through single-gene polymerase chain reaction (PCR)[2] technology. To date, a limited number of microarray experiments using platelet-derived mRNAs have been published; these studies generally agree on platelet transcript quantitation and gene expression patterns.[6,16,17] Furthermore, it is clear that this approach provides an efficient means to identify novel genes and proteins functionally expressed in human platelets.[6] Not surprisingly, platelets retain fewer transcripts—ranging from approximately 1600 to 3000 mRNAs—than those found in nucleated cells[6,17,18] (Table 1). This limited number of platelet-expressed transcripts presumably represents the lack of ongoing transcription in the anucleate platelet.

Given that most microarray protocols are semiquantitative, microarray findings must be validated by other techniques such as quantitative polymerase chain reaction (Q-PCR). The combination of 2 complementary transcript profiling techniques, microarray and SAGE, demonstrated that 50% to 89% of platelet RNA tags are mitochondrial transcripts presumably related to persistent mitochondrial transcription in the absence of nuclear-derived transcripts.[19] The overrepresentation of mitochondrial transcripts in platelets is considerably greater than that of its closest cell type (skeletal muscle), of which 20% to 25% are mitochondrial SAGE tags.[20] Thus, although SAGE clearly has advantages in cellular transcript profiling, its applicability in platelet diseases appears limited. Nonetheless, the relative enrichment of mitochondria-derived transcripts does not interfere with platelet microarray studies.

The feasibility of analyzing platelet transcripts from a single platelet donor with as little as 50 ng total platelet RNA, or approximately 40 mL whole blood, has been demonstrated.[16] Reliable mRNA amplification was validated by Q-PCR and by parallel hybridization of amplified and nonamplified RNA samples. In this study, gene profiling results were reproducible for 9815 of 9850 represented genes, providing initial proof that this approach can be applied in platelet-related human diagnostic studies starting from small sample volumes. Recently, microarray analysis was used to identify genes that are differentially expressed in several platelet-related diseases (see "Global profiling to study platelet-associated disorders," below). These studies clearly establish the feasibility of platelet transcript profiling in identifying differentially expressed genes, characterizing novel platelet-expressed genes, and elucidating the molecular signature of a disease with potential application for platelet diagnostics.

MicroRNA profiling of platelets

MicroRNAs (miRNAs) are a highly conserved class of short, noncoding RNAs that regulate gene expression during cell differentiation, proliferation, and apoptosis. To dissect regulatory pathways that control megakaryocytic differentiation, miRNA expression profiling was performed on in vitro–differentiated megakaryocytes derived from $CD34^+$ hematopoietic progenitors.[21] Several miRNAs were identified, with the subsequent demonstration that miR-130a targets *MAFB*, a transcription factor that is up-regulated during megakaryocytic differentiation and induces the glycoprotein IIb (GPIIB) gene *ITGA2B*. Moreover, the up-regulation of miR-101, miR-126, miR-99a, miR-135, and miR-20 was documented in megakaryoblastic leukemic cell lines compared with in vitro–differentiated megakaryocytes and $CD34^+$ progenitors. These data suggest an important regulatory role of miRNAs during megakaryocytopoiesis; however, a role for miRNAs in thrombopoiesis remains unestablished.

Limitations and perspectives of transcript profiling

Major limitations of modern transcript profiling approaches include reliable and reproducible detection of low-abundant transcripts, feasibility of truly quantitative transcript profiling, bulky and complex data processing, and (in the case of platelets) limited amounts of RNA. Furthermore, accurate platelet transcript profiling requires stringent attention to purification methodologies because a single nucleated cell (ie, leukocyte) contains considerably more mRNA than a platelet.[22] It has become evident that the platelet transcriptome is complex and dynamically controlled at different levels, including regulation by miRNAs,[21] signal-dependent pre-mRNA splicing,[5] and translational control pathways such as mammalian target of rapamycin (mTOR).[4] Despite significant progress in microarray chip design, accurate transcript profiling still requires validation by Q-PCR or other techniques. Efficient mRNA amplification,[16] development of more sensitive whole genome microarrays (which detect alternatively spliced transcripts),[9] and enhancements to bioinformatics software should obviate some of these restrictions.

Platelet protein analyses: proteomics

Modern proteomic techniques

The proteome is the full set of proteins expressed by a genome under a particular set of environmental conditions.[23] The strategy for platelet proteomic analysis generally incorporates platelet fractionation, protein separation, and tryptic digestion methodologies followed by protein identification (Figure 1). Improvements in each of these steps enhance the sensitivity and accuracy of protein identification. Proteomic experiments may begin with protein separation (either in-gel or non-gel); alternatively, the entire mixture of protein can be digested before protein identification (for reviews, see Perrotta and Bahou,[24] de Hoog and Mann,[25] Aebersold and Mann,[26] and Steen and Mann[27]). The availability of the human genome sequence coupled with advances in bioinformatics, computer technology, and mass spectrometry (MS) provide for large-scale, robust, and automated analyses of the cellular proteome.

Table 1. Platelet microarray studies

Reference	No. genes studied	No. present/no. marginal	No. present	No. arrays	Platelet source
Gnatenko et al[6]	12 599	1500/2147	NR	3	Apheresis
Sauer et al[101]	22 200	~1668/3562*	NR	11*	Apheresis
Bugert et al[18]	9 850	NR	~1526	6	Concentrates
McRedmond et al[17]	12 599	NR	2928	1 (in duplicate) from 23 pooled donors	Blood (50 mL)

Numbers of genes studied were the numbers of probes (genes) represented on individual microarray slides.
NR indicates not reported.
*Includes normal and essential thrombocythemic platelets.

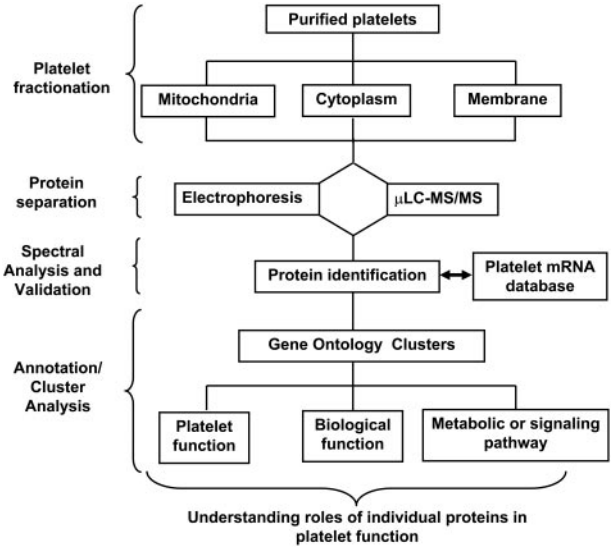


Figure 1. Schema outlining a general approach to platelet proteomic analysis. An ideal experiment generally incorporates platelet subfractionation methods with sophisticated mass spectrometric techniques and computational analyses to elucidate platelet biological functions.

Several mass spectrometry technologies have been developed, including electrospray ionization (ESI-MS),[28] matrix-assisted laser desorption/ionization (MALDI)[29,30] with variations such as ESI-TOF (time of flight) and MALDI-TOF, liquid chromatography coupled to tandem mass spectrometry (μLC-MS/MS), and multidimensional protein identification technology (MudPIT).[31,32] Self-assembled monolayers (SAMs) using trypsin attached to gold have been used to identify 107 platelet proteins, thus establishing the feasibility of proteomic biosensors.[33] More recently, several techniques have been devised to directly compare cellular proteomes, including isotope-coded affinity tags (ICATs)[34] and isobaric tags (iTRAQ)[35-37] coupled to μLC-MS/MS.

Quantitative proteomic[38] techniques are used not only to estimate protein abundance but also to compare cell proteomes. These strategies involve labeling peptides derived from individual samples with isobaric tags (ie, same mass, distinct tags) that fragment into different reporter ions on collision-induced dissociation (CID). Peak area ratios of the reporter ions are used to quantify relative peptide abundance.[39] iTRAQ technology is potentially superior to ICAT-labeling strategies, which detect only cysteine-containing peptides and typically discard peptides that have been posttranslationally modified.[34,39] Other quantitative techniques, such as carboxyl termini labeling of platelet-derived tryptic peptides with oxygen-18, have been used to study the differential expression of platelet proteins.[40]

Limitations of proteomic techniques: sensitivity and the dynamic range

The platelet proteome is subject to rapid changes in response to external signals. Accordingly, accurate and comprehensive platelet protein profiling requires strict attention to technical details to ensure reproducibility and to minimize protein losses resulting from proteolysis, sample preparation, and protein isolation.

Modern proteomic approaches are limited by large differences in the concentrations of the most and least abundant cellular proteins (approximately 5-log difference). Because data acquisition using tandem MS (MS/MS) is triggered by ion abundance levels,[41] peptides derived from more abundant proteins may obscure those derived from less abundant proteins. Thus, this technique can be biased against low-abundance ion signals. In yeast, the dynamic range is insufficient to effectively sample low-abundance proteins of less than 100 copies/cell.[42] Other technical limitations in proteomic techniques include detection of proteins with extremes in pI and molecular weight[43] and membrane-associated or -bound proteins.[44]

Accurate platelet protein profiling requires that platelets be stringently purified of contaminating erythrocytes, leukocytes, and plasma proteins. Table 2 summarizes typical platelet RNA and protein yields obtained from 10 mL peripheral blood, adapting a 2-step purification procedure that incorporates centrifugation, gel filtration, and immunodepletion.[6] Although the amounts of total platelet RNA (approximately 1 μg) and protein are sufficient for most experiments, detection of low-abundance proteins cannot be enhanced because methods for protein amplification (unlike gene-based PCR technologies) are nonexistent.

This issue of protein sensitivity is further highlighted in Table 3, which delineates platelet RNA and protein abundance distributions through composite microarray studies[6] or results from a representative platelet proteomic experiment with μLC-MS/MS spectral counts as estimates of relative protein abundance.[41] Although this distribution of platelet transcript abundances follows a near-normal distribution, the relative protein abundances are distributed differently. Most detected proteins are identified with 1, 2, or 3 peptide hits per protein. Indeed, these results are in agreement with another study in which 62.6% of proteins were identified by a single peptide, with most of the peptide hits corresponding to highly abundant platelet proteins.[45] These data suggest that platelet protein identification using available proteomic technologies has limited ability to quantify peptides with low spectral counts[17] and that protein identification is skewed toward more abundant cellular proteins.[46,47] Although the number of "one-hit wonders" may potentially be reduced by the better separation of proteins before MS identification, it is unclear whether these limitations can be addressed through other quantitative proteomic approaches.[34-37]

Table 2. Total RNA and protein yields from peripheral blood

	Platelet yields, × 10^9				
	Blood, 10 mL	PRP*	Purification†	Yield	Sensitivity, copies/cell
Total RNA	3.1	2.1	1.0	1 μg	<0.1‡
Protein	—	—	—	2 mg	12§

— indicates not applicable.
*Centrifugation step (platelet-rich plasma isolated by centrifugation).
†Gel filtration with or without $CD45^+$ leukocyte depletion.
‡Calculated for microarray analysis after amplification step.
§Calculated based on silver stain detection limits of 1 ng for a 50-kDa protein.

Table 3. Distribution of transcript and protein abundance

Relative abundance, decile	Transcripts, %	Proteins, %	Average no. of peptide hits per protein
1	20.6	85.1	5.8
2	28.4	7.8	39.2
3	22.0	2.8	66.1
4	9.9	2.1	95.8
5	10.6	1.4	111.0
6	2.8	0	—
7	1.4	0	—
8	2.1	0	—
9	1.4	0	—
10	0.7	0.7	257

In this prototype experiment, solubilized platelet proteins were analyzed by μLC-MS/MS, and protein abundance was quantified from the number of peptide hits per individual protein (in this experiment, peptide hits per protein ranged from 1 to 257). Relative protein abundance was determined by the number of peptide hits per protein, and rank-ordered proteins were assigned to deciles based on abundance (1, least abundant; 10, most abundant). The percentage of proteins per decile and the average number of peptide hits per protein are also delineated. Note that most proteins (>85%) were skewed toward those of low abundance, represented by a small number of peptide hits, though transcript expression is distributed more evenly (abundance of corresponding transcripts was determined as previously reported[90]).

— indicates not applicable.

Preanalytical variability in platelet proteomics

The methods used to isolate and prepare protein samples affect the results of platelet proteomic studies. For example, findings may vary when platelet proteins are precipitated with ethanol or trichloroacetic acid.[48] Delays in sample processing and type of anticoagulant used to collect blood samples may also influence the results of MS identification of platelet proteins.[49] Similarly, studies of the low-molecular–weight (less than 15-kDa) peptidome demonstrate that early and gentle platelet separation is crucial for obtaining reproducible identification of platelet-released peptides and proteins.[50] Other studies have emphasized the importance of protease inhibitors and uniform sample storage in improving data reproducibility.[51] These studies demonstrate a clear requirement for standardization of preanalytical variables to ensure comprehensive, reproducible, and comparative (laboratory-to-laboratory) platelet proteomic analyses.

Platelet proteomic analyses

Platelet proteomic studies can be grouped into 2 distinct yet overlapping subcategories, proteomic analyses of quiescent platelets (the static platelet proteome) or activated platelets (the functional platelet proteome) (Figure 2). Whole proteomic strategies identified many platelet proteins, although subsequent approaches to dissect the changes that occur with platelet activation focused on platelet fractions (membrane proteins) or functional end points (eg, phosphorylation patterns, microparticles) in response to external stimuli. Such subproteomic studies use the same technologies but provide a more detailed analysis of function over time.

Quiescent platelets: the static platelet proteome. Initial studies of the platelet proteome focused on characterizing proteins in resting platelets through combinations of 2-dimensional electrophoresis (2-DE) and in-gel protein detection using monoclonal antibodies.[52-54] Although successful, these approaches were limited by the dynamic range of platelet proteins, coupled with the cost and sensitivity of immunodetection. Similar techniques were subsequently applied to establish a platelet protein map and to characterize tyrosine-phosphorylated proteins in resting platelets.[55] In this study, cytosolic protein fractions were separated by 2-DE, and phosphorylated proteins were immunodetected using antiphosphotyrosine antibodies and MALDI-TOF. One hundred eighty-six protein spectra corresponding to 123 unique proteins were identified through this approach. By making use of immobilized pH gradient (IPG) strips to facilitate the one-dimensional gel separation, 286 platelet proteins were identified in the acidic range (pI range 4-5).[56] A more comprehensive profiling of platelet proteins (pI range 5-11) identified 760 protein features corresponding to 311 different genes, resulting in the annotation of 54% of the 2-DE (pI range 5-11) proteome map.[57]

Combined fractional diagonal chromatography (COFRADIC) is a non-gel–based technique in which peptide sets are sorted in a diagonal reverse-phase chromatography system through a specific modification of their side chains.[58,59] This technique was used to identify 264 platelet proteins present in cytosolic and membrane fractions; the dynamic range spanned 4 to 5 orders of magnitude of protein concentration. Modifications of this technology identified a core set of 641 platelet proteins,[45] the largest platelet protein set reported to date. These proteins were

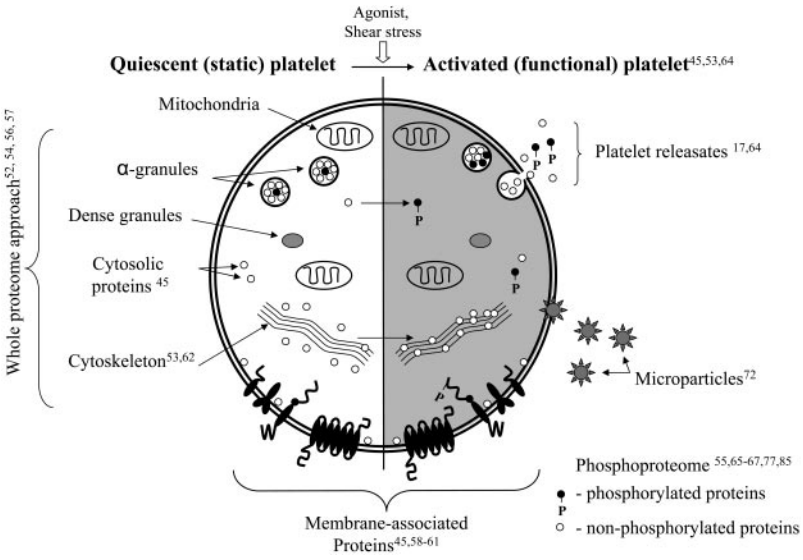


Figure 2. Schema of platelet ultrastructure integrated with proteomic studies. The unshaded panel delineates the proteomic studies involving quiescent platelets. The shaded panel represents those studies focusing on activation-dependent platelet end points (eg, microparticles, exosome/releasates). Superscript numbers refer to references.

classified using the Gene Ontology database, revealing that 16% were membrane proteins (see "Membrane proteins," below), and 64% were present in the cytoskeleton, endoplasmic reticulum, mitochondria, cytosol, or Golgi apparatus. Interestingly, up to 20% of proteins were classified as nucleus restricted. Given that platelets are anucleate, these proteins presumably arise from megakaryocytes during thrombopoiesis; their function in platelets is unclear.

Membrane proteins. The study of platelet integral membrane proteins and surface receptors through traditional 2-DE techniques is limited by the low solubility of these proteins, their association with the platelet membrane, their high molecular weight, and the potentially limited number of diagnostic tryptic cleavage fragments. In addition, the presence of highly abundant cytoskeleton proteins such as actin obscures less abundant platelet membrane proteins.

Two groups have adapted proteomic approaches for the study of platelet membrane proteins.[45,60] Although applications using COFRADIC[45] predicted the presence of 87 putative helix-spanning membrane proteins, a more focused analysis of the platelet membrane proteome was pursued by enriching membrane proteins before protein identification through μLC-MS/MS. In this study, 2 distinct solubilization methods were used to reduce the overrepresentation of cytoskeleton proteins, allowing for the aggregate detection of 233 established or putative transmembrane proteins.[60] Finally, a combination of microarray and mass spectrometric techniques has been used to identify novel membrane proteins that signal during platelet aggregation and to characterize proteins that become phosphorylated on tyrosine or serine residues on platelet aggregation.[61]

Activated platelet proteome: activation-dependent platelet complexes. An early study of activation-dependent responses in platelets identified 27 proteins that were translocated to the actin cytoskeleton during platelet activation.[62] Subsequently, a related study that focused on characterizing static protein complexes as a prelude to dissecting activation-dependent complexes was undertaken in accordance with a novel 2-DE technology that combined 1-dimensional separation of monomeric and multimeric proteins in their native state followed by a 2-dimensional denaturing step.[63] In total, 63 platelet proteins were identified, among them the integral membrane proteins GPIbβ, CD9, and CD36 and the fibrinogen receptor $\alpha_{IIb}\beta_3$. Mitochondrial proteins were detected in the membrane fraction, several of which were organized into known complexes.

Proteins secreted during platelet activation. Proteomic strategies have been used to identify protein subsets secreted during platelet activation. Such strategies resulted in the initial identification of 82 secreted proteins[17]; this was followed by a more robust analysis of the platelet secretome.[64] With the use of thrombin-stimulated platelets, a combination of MALDI-TOF and MudPIT identified more than 300 released proteins.[64] Thirty-seven percent of these proteins were previously known to be released from platelets, whereas another 35% were reported to be released from other secretory cells. The remaining 28% of proteins were not known to be released by any cell type. Moreover, several of the secreted proteins have been previously identified in human atherosclerotic lesions but absent in normal vasculature. In this study, protein profiling resulted in the identification of potential targets for future drug development (ie, secretogranin III, cyclophilin A, and calumenin) and shed light on possible mechanisms of platelet adhesion that contribute to the development of atherosclerosis.

Platelet phosphoproteome. Tyrosine phosphorylation plays a key role in the activation-dependent regulation of protein function, signal transduction, and other complex cellular processes. Several studies have used phosphotyrosine-dependent antibodies and proteomic technologies to dissect intracellular signaling events that occur with thrombin activation, consistently demonstrating an ability to identify discrete proteins, subsets of which had been previously undescribed.[65,66] By using the thrombin receptor activating peptide (TRAP) to specifically activate proteinase-activated receptor 1 (PAR1), 62 differentially phosphorylated proteins were detected, 41 of which were identified by μLC-MS/MS.[67] Interestingly, 8 of these had never been described, and the protein repertoire was shown to originate from 31 genes, highlighting how alternative splicing expands the platelet proteome. Although these initial studies established the feasibility of using proteomic approaches to dissect ligand-dependent phosphorylation changes,[68,69] it is likely that technical advances will further enhance its robustness.[70]

Platelet microparticles. When platelets are activated in vivo, 2 types of membrane vesicles are released, microparticles (which bud from the plasma membrane) and exosomes.[71] Microparticles range in size from 0.1 to 1.0 μm, whereas exosomes are smaller and range from 40 to 100 nm. Platelet microparticles are abundant, exert procoagulant activity, and play hemostatically critical roles in several clinical disorders, including heparin-induced thrombocytopenia and immune thrombocytopenic purpura. A recent proteomic study provided the first panoramic overview of microparticles, identifying 578 proteins that constitute this subcellular proteome.[72] As expected, many of these represented well-characterized platelet proteins. Surprisingly, 380 of 578 proteins had not been previously described in platelet proteomic studies, suggesting these platelet fragments have a unique protein composition.

Integrating platelet transcriptomic and proteomic studies

Recent evidence from mathematical modeling studies demonstrates the need to delineate mRNA and protein expression levels to optimally map intracellular networks. When applied to nucleated cells, these integrated platforms demonstrate some correlation between transcript and protein expression.[73] In platelets, this relationship is even more complex because platelets lack a genome and ongoing RNA transcription, display minimal translational activity, have the capacity for signal-dependent translation of selected mRNAs[4] and inducible (activation-dependent) transcript splicing,[5] absorb select plasma proteins, and contain residual megakaryocyte-derived mRNAs and proteins.

Initial attempts to correlate platelet mRNA and protein profiles have been described.[17,74] These studies demonstrate that up to 69% of secreted and cytosolic proteins were detectable at the mRNA level, suggesting relatively good correlation between proteomic and transcriptomic data in the study of end points of detection and identification.[17] The authors concluded that despite the absence of gene transcription, the platelet proteome is mirrored in the transcriptome, and transcriptional analysis predicts the presence of novel proteins in the platelet. The lack of detailed quantitative correlations limits the ability to compare platelet gene and protein levels with those of nucleated cells.[73]

Global profiling to study platelet-associated disorders

Platelet proteomic studies

Glanzmann thrombasthenia. GT is an autosomal recessive disease characterized by the complete absence or a marked reduction in the $\alpha_{IIb}\beta_3$ heterodimeric complex.[75,76] Compared with normal platelets, GT platelets exhibit a low tyrosine phosphorylation profile, confirming the key role of functional $\alpha_{IIb}\beta_3$ in the initiation of protein tyrosine phosphorylation.[77] Several proteins displayed

attenuated thrombin-dependent phosphorylation kinetics despite normal initial phosphorylation rates. Similar results were obtained by inhibiting thrombin aggregation of control platelets with $\alpha_{IIb}\beta_3$ antagonists or in the absence of stirring. These results suggest that tyrosine phosphorylation of specific proteins is dependent on thrombin activation during early and late steps of $\alpha_{IIb}\beta_3$ engagement in aggregation. No other clear differences were identified between normal and GT platelets.

Ischemic stroke. Genetic risk factors have been identified in patients with thrombophilia, though associations are the strongest for patients with venous thromboembolic disease.[78] Thus, whereas hyperhomocysteinemia and antiphospholipid/anticardiolipin antibodies are stroke risk factors,[79] congenital thrombophilic states (factor V Leiden and prothrombin gene mutations; protein C, protein S, and antithrombin III deficiencies) account for a disproportionately small subset of ischemic stroke,[80] with the greatest risk in younger persons.[81] In contrast, differential expression of platelet proteins may favor platelet activation and thrombus formation. The best evidence in support of this observation stems from correlative studies involving platelet cell surface glycoprotein receptors such as $\alpha_2\beta_1$, in which higher expression levels are associated with increased collagen binding,[82] and greater risk for ischemic heart disease in homozygotes, especially smokers.[83] Similarly, platelet membrane polymorphisms have been linked to stroke in small studies, but the evidence is not strong, and it is most significant in young patients.[84]

A phosphoproteomic comparison of platelets from healthy controls and stroke patients was completed using antiphosphotyrosine antibodies.[85] In unstimulated platelets, a discrete subset of proteins displayed significantly greater tyrosine phosphorylation in 85% of the 20 stroke patients studied than in the healthy controls. Additionally, the authors identified other tyrosine-phosphorylated bands in the stroke platelets that were absent in the resting platelets of the control patients. Although these results may have some prognostic merit in monitoring patients with cerebrovascular insufficiency,[85] they provide no specific protein identification and fail to distinguish cause from effect in dissecting proteomic differences that pathogenetically define platelet-associated stroke risk.

Toxicology. Benzene exposure is an established risk factor for acute myeloid leukemia and may play a role in other human blood diseases.[86] Surface-enhanced laser desorption ionization-TOF MS (SELDI-TOF MS) was used to compare the blood serum proteome of 40 shoe factory workers who experienced well-characterized occupational exposure to benzene with the blood serum proteome of unexposed controls to identify potential biomarkers of benzene exposure.[87] Three low-molecular–weight proteins were consistently down-regulated in exposed subjects. Two of these proteins were identified as platelet-derived CXC chemokines, platelet factor 4 (PF4), and connective tissue–activating peptide III (CTAP III). These findings suggest that lower expression of these proteins may serve as a potential biomarker of early benzene toxicity.

From the transcriptome to the proteome

With today's technology, it is feasible to adapt transcriptomic analysis to predict differences in the proteome, to validate these differences between normal and diseased platelets, and to exploit these differences to develop modern diagnostic tests.[19,74,88]

Direct platelet transcriptomics: essential thrombocythemia as a paradigm. The most compelling evidence that transcript profiling can distinguish diseased from normal platelet profiles is based on studies of essential thrombocythemia (ET) that were conducted with megakaryocytes and platelets.[89,90] ET is a myeloproliferative disorder characterized by increased proliferation of megakaryocytes, elevated numbers of morphologically normal circulating platelets, and considerable thrombohemorrhagic events.[91] Microarray profiling has been used to analyze platelet transcripts of 6 patients with ET and 5 healthy controls.[90] Initial analysis demonstrated different molecular signatures capable of distinguishing normal platelets from ET platelets. Moreover, ET platelets collectively expressed higher numbers of individual transcripts than normal platelets but considerably less distinct transcripts than nucleated cells.[6] Statistical analysis revealed 170 genes that were differentially expressed between ET and normal platelets, most (141 of 170 genes) of which were up-regulated in ET platelets.

Among differentially expressed genes, the transcript encoding 17β-hydroxysteroid dehydrogenase type 3 (*HSD17B3*) was dramatically reduced in ET platelets. This enzyme, not previously characterized in human platelets, belongs to a larger family of *17BHSD*s that play a role in normal and abnormal testosterone biosynthesis.[92] Transcripts for 2 other *17BHSD* family members were identified in platelets; they are *HSD17B11* and *HSD17B12*. Absence of *HSD17B3* transcript expression was evident in all 6 patients with ET. These changes occurred concomitantly with elevated transcript levels of *HSD17B12* in the same patient subgroup. In contrast, the expression of *HSD17B3* in normal platelets was accompanied by negligible *HSD17B12* expression. At the protein level, HSD17β3 enzyme activity was demonstrated in platelet lysates, confirming that human platelets retained the capacity to catalyze the final step in gonadal testosterone synthesis. These data demonstrate the potential of using transcriptomics to lend focus to proteomic studies.

Indirect transcriptomic approaches of platelet disorders. *Gray platelet syndrome.* Gray platelet syndrome (GPS) is a rare platelet disorder manifest by bleeding, thrombocytopenia, and a distinct lack of α-granules.[93,94] Recently, microarray analysis was used to study molecular mechanisms involved in GPS, focusing on fibroblasts because these cells may be involved in the transition to myelofibrosis evident in subsets of patients with the disorder.[95] By comparing microarray profiles of normal and GPS fibroblasts, the up-regulation of various proteins (fibronectin 1, thrombospondins 1 and 2, and collagen VIα) was demonstrated. Overexpression of fibronectin and thrombospondin 1 was confirmed at the RNA level by Northern blot analysis, whereas fibronectin overexpression was confirmed at the protein level by immunostaining. Although this study fails to address the role of megakaryocyte-derived proteins in the development of myelofibrosis, it does suggest that distinct protein subsets may play a role in the progression to myelofibrosis seen in some GPS patients.

Polycythemia rubra vera. Like ET, polycythemia rubra vera (PRV) is a myeloproliferative disorder associated with thrombotic complications.[96] Although the molecular basis for enhanced thrombotic risk remains unknown, defects affecting the platelet thrombohemorrhagic balance are envisioned; however, ancillary leukocyte dysfunction linking the inflammatory response to risk for thrombosis cannot be excluded. To date, microarray studies have been restricted to PRV leukocytes to the exclusion of platelets.[97-99] Furthermore, limited correlative proteomic studies validate the differences seen by microarray or stratify transcriptomic differences by thrombotic phenotype. Although further correlative studies are needed, PRV platelets, like ET platelets, may have a distinct proteomic profile that could help explain the pathogenesis of thrombosis in this hematopoietic disorder.

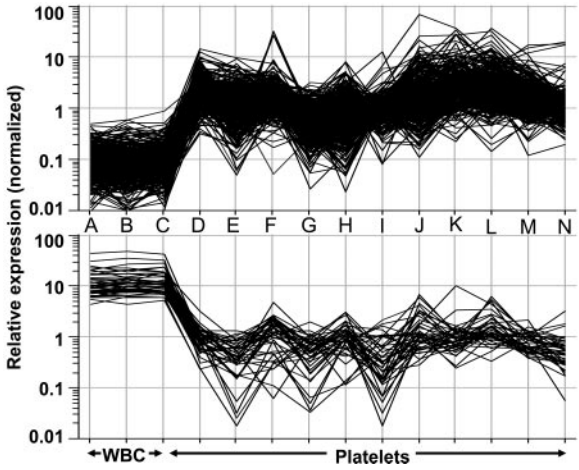


Figure 3. In silico gene rank–intensity plots from a first-generation platelet gene chip. Normalized data from individual microarray analyses obtained from 3 distinct leukocyte samples (A-C), 5 normal platelet samples (D-H), and 6 essential thrombocythemic platelet samples (I-N)[90] were analyzed by one-way ANOVA using parametric testing to identify a 432-member gene list. (top) Relative expression of the platelet-restricted genes (n = 389). (bottom) Expression intensity of the leukocyte-restricted genes (n = 43). Note the clear difference in the expression patterns between the 2 groups (leukocyte vs platelet).

Future directions and applicability

Custom platelet chips

The expense of global transcript and platelet protein profiling precludes investigators from performing larger disease cohort studies. The development of cost-effective and reliable tools for transcript profiling would facilitate platelet-related diagnostics and experimentation. Thus, we have fabricated a custom platelet microarray that can be used in conjunction with proteomic analyses to study normal and diseased platelets. The final transcript list was established by analyzing microarray results from normal and ET platelets[90] and included platelet-restricted genes without detectable expression in leukocytes, genes that appear to be associated with a thrombohemorrhagic ET phenotype, and genes whose platelet expression is more than 10-fold greater than its leukocyte expression or whose leukocyte expression is more than 10-fold greater than its platelet expression. Several *Arabidopsis* probe elements were included to serve as normalization controls and to minimize interslide and intraslide variability.[100] After removing duplicates, the final list contains 432 genes that clearly cosegregate by cell type (Figure 3). Although initial analyses demonstrate the potential of this chip to distinguish ET platelets from normal platelets (data not shown), further studies are ongoing. Nonetheless, the development of platelet-specific gene and protein chips could lead to more widespread applicability of this technology to platelet-related and thrombotic disorders.

General profiling perspectives

Future advances in proteomic technology that incorporate miniaturization,[101] coupled with an ability to integrate functional genomics and proteomics,[102] will help unravel the complex biological pathways that are relevant to platelet-associated disorders. The human genome encodes between 20 000 and 40 000 genes, whereas the estimated number of functional proteins may number as much as half a million because of alternative splicing, translational regulation, and posttranslational modifications.[102] Because of this large number of functional proteins, we anticipate that integrated analysis of the transcriptome and the proteome are required to optimally dissect the molecular mechanisms responsible for platelet-related diseases (Figure 4). Parallel advances in quantitative proteomic techniques will also have an impact on the field,[34-37] resulting in the identification of platelet-related disease biomarkers and novel therapeutic targets.

Acknowledgments

This work was supported by NIH grants HL49141, HL53665, HL76457; Department of Defense grant MPO48005; a Targeted Research Award from Stony Brook University; and National Institutes of Health Center grant MO1 10710-5 to the University Hospital General Clinical Research Center.

Authorship

Contribution: D.V.G., P.L.P, and W.F.B. reviewed the literature and wrote the paper.

Conflict-of-interest disclosure: The authors declare no competing financial interests.

Correspondence: Dmitri V. Gnatenko, Department of Medicine, Division of Hematology, T15-040 HSC, State University of New York, Stony Brook, NY 11794-8151; e-mail: dgnatenko@notes.cc.sunysb.edu.

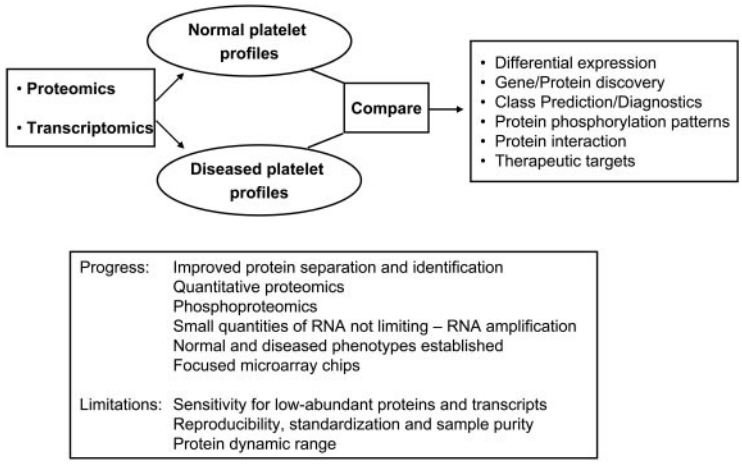


Figure 4. Integration of platelet transcript and protein profiling to study human diseases. The combination of proteomic and transcriptomic technologies can be applied for comparative studies between normal and diseased platelets, ultimately leading to novel diagnostic assays or to the identification of novel therapeutic targets. Potential applications include treatment of not only single-gene platelet disorders but also the broader subset of patients with platelet-related cardiovascular or cerebrovascular disease. The box summarizes current limitations and progress achieved to date.

References

1. Stenberg PE, Hill RL. Platelets and megakaryocytes. In: Lee G, Foerster J, Lukens J, eds. Wintrobe's Clinical Hematology. Philadelphia, PA: Lippincott Williams & Wilkins; 1999:615-660.
2. Newman PJ, Gorski J, White GC 2nd, Gidwitz S, Cretney CJ, Aster RH. Enzymatic amplification of platelet-specific messenger RNA using the polymerase chain reaction. J Clin Invest. 1988;82:739-743.
3. Kieffer N, Guichard J, Farcet JP, Vainchenker W, Breton-Gorius J. Biosynthesis of major platelet proteins in human blood platelets. Eur J Biochem. 1987;164:189-195.
4. Weyrich AS, Dixon DA, Pabla R, et al. Signal-dependent translation of a regulatory protein, Bcl-3, in activated human platelets. Proc Natl Acad Sci U S A. 1998;95:5556-5561.
5. Denis MM, Tolley ND, Bunting M, et al. Escaping the nuclear confines: signal-dependent pre-mRNA splicing in anucleate platelets. Cell. 2005;122:379-391.
6. Gnatenko DV, Dunn JJ, McCorkle SR, Weissmann D, Perrotta PL, Bahou WF. Transcript profiling of human platelets using microarray and serial analysis of gene expression. Blood. 2003;101:2285-2293.
7. Hoheisel JD. Microarray technology: beyond transcript profiling and genotype analysis. Nat Rev Genet. 2006;7:200-210.
8. Sausville EA, Holbeck SL. Transcription profiling of gene expression in drug discovery and development: the NCI experience. Eur J Cancer. 2004;40:2544-2549.
9. Butte A. The use and analysis of microarray data. Nat Rev Drug Discov. 2002;1:951-960.
10. Velculescu V, Zhang L, Vogelstein B, Kinzler K. Serial analysis of gene expression. Science. 1995;270:484-487.
11. Zhang L, Zhou W, Velculescu V, et al. Gene expression profiles in normal and cancer cells. Science. 1997;276:1268-1272.
12. Dunn JJ, McCorkle SR, Praissman LA, et al. Genomic signature tags (GSTs): a system for profiling genomic DNA. Genome Res. 2002;12:1756-1765.
13. Wang E, Miller L, Ohnmacht G, Liu E, Marincola F. High-fidelity mRNA amplification for gene profiling. Nat Biotechnol. 2000;18:157-459.
14. Peters DG, Kassam AB, Yonas H, O'Hare EH, Ferrell RE, Brufsky AM. Comprehensive transcript analysis in small quantities of mRNA by SAGE-lite. Nucleic Acids Res. 1999;27:e39.
15. Wicki AN, Walz A, Gerber-Huber SN, Wenger RH, Vornhagen R, Clemetson KJ. Isolation and characterization of human blood platelet mRNA and construction of a cDNA library in lambda gt11: confirmation of the platelet derivation by identification of GPIb coding mRNA and cloning of a GPIb coding cDNA insert. Thromb Haemost. 1989;61:448-453.
16. Rox JM, Bugert P, Muller J, et al. Gene expression analysis in platelets from a single donor: evaluation of a PCR-based amplification technique. Clin Chem. 2004;50:2271-2278.
17. McRedmond JP, Park SD, Reilly DF, et al. Integration of proteomics and genomics in platelets: a profile of platelet proteins and platelet-specific genes. Mol Cell Proteomics. 2004;3:133-144.
18. Bugert P, Dugrillon A, Gunaydin A, Eichler H, Kluter H. Messenger RNA profiling of human platelets by microarray hybridization. Thromb Haemost. 2003;90:738-748.
19. Dittrich M, Birschmann I, Stuhlfelder C, et al. Understanding platelets: lessons from proteomics, genomics and promises from network analysis. Thromb Haemost. 2005;94:916-925.
20. Welle S, Bhatt K, Thornton C. Inventory of high-abundance mRNAs in skeletal muscle of normal men. Genome Res. 1999;9:506-513.
21. Garzon R, Pichiorri F, Palumbo T, et al. MicroRNA fingerprints during human megakaryocytopoiesis. Proc Natl Acad Sci U S A. 2006;103:5078-5083.
22. Bahou WF, Gnatenko DV. Platelet transcriptome: the application of microarray analysis to platelets. Semin Thromb Hemost. 2004;30:473-484.
23. Pandey A, Mann M. Proteomics to study genes and genomes. Nature. 2000;405:837-846.
24. Perrotta PL, Bahou WF. Proteomics in platelet science. Curr Hematol Rep. 2004;3:462-469.
25. de Hoog CL, Mann M. Proteomics. Annu Rev Genomics Hum Genet. 2004;5:267-293.
26. Aebersold R, Mann M. Mass spectrometry-based proteomics. Nature. 2003;422:198-207.
27. Steen H, Mann M. The ABC's (and XYZ's) of peptide sequencing. Nat Rev Mol Cell Biol. 2004;5:699-711.
28. Fenn JB, Mann M, Meng CK, Wong SF, Whitehouse CM. Electrospray ionization for mass spectrometry of large biomolecules. Science. 1989;246:64-71.
29. Hillenkamp F, Karas M, Beavis RC, Chait BT. Matrix-assisted laser desorption/ionization mass spectrometry of biopolymers. Anal Chem. 1991;63:1193A-1203A.
30. Hillenkamp F, Karas M. Mass spectrometry of peptides and proteins by matrix-assisted ultraviolet laser desorption/ionization. Methods Enzymol. 1990;193:280-295.
31. Washburn MP, Wolters D, Yates JR 3rd. Large-scale analysis of the yeast proteome by multidimensional protein identification technology. Nat Biotechnol. 2001;19:242-247.
32. Link AJ, Eng J, Schieltz DM, et al. Direct analysis of protein complexes using mass spectrometry. Nat Biotechnol. 1999;17:676-682.
33. Tyan YC, Liao JD, Jong SB, et al. Proteomic profiling of platelet proteins by trypsin immobilized self-assembled monolayers digestion chip and protein identification using electrospray ionization tandem mass spectrometry. J Biomed Mater Res A. 2004;71:90-97.
34. Gygi S, Rist B, Gerber S, Turecek F, Gelb M, Aebersold R. Quantitative analysis of complex protein mixtures using isotope-coded affinity tags. Nat Biotechnol. 1999;17:994-999.
35. Blagoev B, Ong SE, Kratchmarova I, Mann M. Temporal analysis of phosphotyrosine-dependent signaling networks by quantitative proteomics. Nat Biotechnol. 2004;22:1139-1145.
36. Choe LH, Aggarwal K, Franck Z, Lee KH. A comparison of the consistency of proteome quantitation using two-dimensional electrophoresis and shotgun isobaric tagging in *Escherichia coli* cells. Electrophoresis. 2005;26:2437-2449.
37. Ross PL, Huang YN, Marchese JN, et al. Multiplexed protein quantitation in *Saccharomyces cerevisiae* using amine-reactive isobaric tagging reagents. Mol Cell Proteomics. 2004;3:1154-1169.
38. Mann M. Quantitative proteomics? Nat Biotechnol. 1999;17:954-955.
39. Han D, Eng J, Zhou H, Aebersold R. Quantitative profiling of differentiation-induced microsomal proteins using isotope-coded affinity tags and mass spectrometry. Nat Biotechnol. 2001;19:946-951.
40. Staes A, Demol H, Van Damme J, Martens L, Vandekerckhove J, Gevaert K. Global differential non-gel proteomics by quantitative and stable labeling of tryptic peptides with oxygen-18. J Proteome Res. 2004;3:786-791.
41. Liu H, Sadygov RG, Yates JR 3rd. A model for random sampling and estimation of relative protein abundance in shotgun proteomics. Anal Chem. 2004;76:4193-4201.
42. Ghaemmaghami S, Huh WK, Bower K, et al. Global analysis of protein expression in yeast. Nature. 2003;425:737-741.
43. Gygi S, Rochon Y, Franza B, Aebersold R. Correlation between protein and mRNA abundance in yeast. Mol Cell Biol. 1999;19:1720-1730.
44. Molloy M. Two-dimensional electrophoresis of membrane proteins using immobilized pH gradients. Anal Biochem. 2000;280:1-10.
45. Martens L, Van Damme P, Van Damme J, et al. The human platelet proteome mapped by peptide-centric proteomics: a functional protein profile. Proteomics. 2005;5:3193-3204.
46. Veenstra TD, Conrads TP, Issaq HJ. What to do with one-hit wonders? Electrophoresis. 2004;25:1278-1279.
47. Watson SP, Bahou WF, Fitzgerald D, Ouwehand W, Rao AK, Leavitt AD. Mapping the platelet proteome: a report of the ISTH Platelet Physiology Subcommittee. J Thromb Haemost. 2005;3:2098-2101.
48. Zellner M, Winkler W, Hayden H, et al. Quantitative validation of different protein precipitation methods in proteome analysis of blood platelets. Electrophoresis. 2005;26:2481-2489.
49. Banks RE, Stanley AJ, Cairns DA, et al. Influences of blood sample processing on low-molecular-weight proteome identified by surface-enhanced laser desorption/ionization mass spectrometry. Clin Chem. 2005;51:1637-1649.
50. Tammen H, Schulte I, Hess R, et al. Peptidomic analysis of human blood specimens: comparison between plasma specimens and serum by differential peptide display. Proteomics. 2005;5:3414-3422.
51. Rai AJ, Gelfand CA, Haywood BC, et al. HUPO Plasma Proteome Project specimen collection and handling: towards the standardization of parameters for plasma proteome samples. Proteomics. 2005;5:3262-3277.
52. Hanash SM, Neel JV, Baier LJ, Rosenblum BB, Niezgoda W, Markel D. Genetic analysis of thirty-three platelet polypeptides detected in two-dimensional polyacrylamide gels. Am J Hum Genet. 1986;38:352-360.
53. Giometti CS, Anderson NG. Protein changes in activated human platelets. Clin Chem. 1984;30:2078-2083.
54. Gravel P, Sanchez JC, Walzer C, et al. Human blood platelet protein map established by two-dimensional polyacrylamide gel electrophoresis. Electrophoresis. 1995;16:1152-1159.
55. Marcus K, Immler D, Sternberger J, Meyer H. Identification of platelet proteins separated by two-dimensional gel electrophoresis and analyzed by matrix assisted laser desorption/ionization-time of flight-mass spectrometry and detection of tyrosine-phosphorylated proteins. Electrophoresis. 2000;21:2622-2636.
56. O'Neill EE, Brock CJ, von Kriegsheim AF, et al. Towards complete analysis of the platelet proteome. Proteomics. 2002;2:288-305.
57. Garcia A, Prabhakar S, Brock CJ, et al. Extensive analysis of the human platelet proteome by two-dimensional gel electrophoresis and mass spectrometry. Proteomics. 2004;4:656-668.
58. Gevaert K, Ghesquiere B, Staes A, et al. Reversible labeling of cysteine-containing peptides allows their specific chromatographic isolation for non-gel proteome studies. Proteomics. 2004;4:897-908.
59. Gevaert K, Goethals M, Martens L, et al. Exploring proteomes and analyzing protein processing by mass spectrometric identification of sorted N-terminal peptides. Nat Biotechnol. 2003;21:566-569.

60. Moebius J, Zahedi RP, Lewandrowski U, Berger C, Walter U, Sickmann A. The human platelet membrane proteome reveals several new potential membrane proteins. Mol Cell Proteomics. 2005;4:1754-1761.
61. Nanda N, Bao M, Lin H, et al. Platelet endothelial aggregation receptor 1 (PEAR1), a novel epidermal growth factor repeat-containing transmembrane receptor, participates in platelet contact-induced activation. J Biol Chem. 2005;280: 24680-24689.
62. Gevaert K, Eggermont L, Demol H, Vandekerckhove J. A fast and convenient MALDI-MS based proteomic approach: identification of components scaffolded by the actin cytoskeleton of activated human thrombocytes. J Biotechnol. 2000;78:259-269.
63. Claeys D, Geering K, Meyer BJ. Two-dimensional Blue Native/sodium dodecyl sulfate gel electrophoresis for analysis of multimeric proteins in platelets. Electrophoresis. 2005;26:1189-1199.
64. Coppinger JA, Cagney G, Toomey S, et al. Characterization of the proteins released from activated platelets leads to localization of novel platelet proteins in human atherosclerotic lesions. Blood. 2004;103:2096-2104.
65. Marcus K, Moebius J, Meyer HE. Differential analysis of phosphorylated proteins in resting and thrombin-stimulated human platelets. Anal Bioanal Chem. 2003;376:973-993.
66. Maguire PB, Wynne KJ, Harney DF, O'Donoghue NM, Stephens G, Fitzgerald DJ. Identification of the phosphotyrosine proteome from thrombin activated platelets. Proteomics. 2002;2:642-648.
67. Garcia A, Prabhakar S, Hughan S, et al. Differential proteome analysis of TRAP-activated platelets: involvement of DOK-2 and phosphorylation of RGS proteins. Blood. 2004;103:2088-2095.
68. Maguire PB. Platelet proteomics: identification of potential therapeutic targets. Pathophysiol Haemost Thromb. 2003;33:481-486.
69. Maguire PB, Foy M, Fitzgerald DJ. Using proteomics to identify potential therapeutic targets in platelets. Biochem Soc Trans. 2005;33:409-412.
70. Zhang Y, Wolf-Yadlin A, Ross PL, et al. Time-resolved mass spectrometry of tyrosine phosphorylation sites in the epidermal growth factor receptor signaling network reveals dynamic modules. Mol Cell Proteomics. 2005;4:1240-1250.
71. Nieuwland R, Sturk A. Platelet-derived microparticles. In: Michelson AD, ed. Platelets. London, United Kingdom: Academic Press; 2002:255-266.
72. Garcia BA, Smalley DM, Cho H, Shabanowitz J, Ley K, Hunt DF. The platelet microparticle proteome. J Proteome Res. 2005;4:1516-1521.
73. Hack CJ. Integrated transcriptome and proteome data: the challenges ahead. Brief Funct Genomic Proteomic. 2004;3:212-219.
74. Macaulay IC, Carr P, Gusnanto A, Ouwehand WH, Fitzgerald D, Watkins NA. Platelet genomics and proteomics in human health and disease. J Clin Invest. 2005;115:3370-3377.
75. Newman PJ, Seligsohn U, Lyman S, Coller BS. The molecular genetic basis of Glanzmann thrombasthenia in the Iraqi-Jewish and Arab populations in Israel. Proc Natl Acad Sci U S A. 1991;88:3160-3164.
76. Nurden AT. Inherited abnormalities of platelets. Thromb Haemost. 1999;82:468-480.
77. Rosa JP, Artcanuthurry V, Grelac F, Maclouf J, Caen JP, Levy-Toledano S. Reassessment of protein tyrosine phosphorylation in thrombasthenic platelets: evidence that phosphorylation of cortactin and a 64-kD protein is dependent on thrombin activation and integrin αIIb β3. Blood. 1997;89:4385-4392.
78. Federman DG, Kirsner RS. An update on hypercoagulable disorders. Arch Intern Med. 2001;161: 1051-1056.
79. Eikelboom JW, Hankey GJ, Anand SS, Lofthouse E, Staples N, Baker RI. Association between high homocyst(e)ine and ischemic stroke due to large- and small-artery disease but not other etiologic subtypes of ischemic stroke. Stroke. 2000;31: 1069-1075.
80. Cushman M, Rosendaal FR, Psaty BM, et al. Factor V Leiden is not a risk factor for arterial vascular disease in the elderly: results from the Cardiovascular Health Study. Thromb Haemost. 1998;79:912-915.
81. Madonna P, de Stefano V, Coppola A, et al. Hyperhomocysteinemia and other inherited prothrombotic conditions in young adults with a history of ischemic stroke. Stroke. 2002;33:51-56.
82. Roest M, Sixma J, Wu Y, et al. Platelet adhesion to collagen in healthy volunteers is influenced by variation of both alpha(2) beta(1) density and von Willebrand factor. Blood. 2000;96:1433-1437.
83. Moshfegh K, Wiullemin W, Redondo M, et al. Association of two silent polymorphisms of platelet glycoprotein Ia/IIa receptor with risk of myocardial infarction: a case-control study. Lancet. 1999; 353:351-354.
84. Frank MB, Reiner AP, Schwartz SM, et al. The Kozak sequence polymorphism of platelet glycoprotein Ibα and risk of nonfatal myocardial infarction and nonfatal stroke in young women. Blood. 2001;97:875-879.
85. Srivastava K, Dash D. Altered membrane fluidity and signal transduction in the platelets from patients of thrombotic stroke. Mol Cell Biochem. 2001;224:143-149.
86. Savitz DA, Andrews KW. Review of epidemiologic evidence on benzene and lymphatic and hematopoietic cancers. Am J Ind Med. 1997;31:287-295.
87. Vermeulen R, Lan Q, Zhang L, et al. Decreased levels of CXC-chemokines in serum of benzene-exposed workers identified by array-based proteomics. Proc Natl Acad Sci U S A. 2005;102: 17041-17046.
88. Stratmann B, Tschoepe D. Pathobiology and cell interactions of platelets in diabetes. Diabetes Vasc Dis Res. 2005;2:16-23.
89. Tenedini E, Fagioli ME, Vianelli N, et al. Gene expression profiling of normal and malignant CD34-derived megakaryocytic cells. Blood. 2004;104: 3126-3135.
90. Gnatenko DV, Cupit LD, Huang EC, Dhundale A, Perrotta PL, Bahou WF. Platelets express steroidogenic 17β-hydroxysteroid dehydrogenases: distinct profiles predict the essential thrombocythemic phenotype. Thromb Haemost. 2005;94: 412-421.
91. Nimer SD. Essential thrombocythemia: another "heterogeneous disease" better understood? Blood. 1999;93:415-416.
92. Geissler WM, Davis DL, Wu L, et al. Male pseudohermaphroditism caused by mutations of testicular 17 beta-hydroxysteroid dehydrogenase 3. Nat Genet. 1994;7:34-39.
93. Srivastava PC, Powling MJ, Nokes TJ, Patrick AD, Dawes J, Hardisty RM. Grey platelet syndrome: studies on platelet alpha-granules, lysosomes and defective response to thrombin. Br J Haematol. 1987;65:441-446.
94. Reed GL. Platelet secretory mechanisms. Semin Thromb Hemost. 2004;30:441-450.
95. Hyman T, Huizing M, Blumberg PM, Falik-Zaccai TC, Anikster Y, Gahl WA. Use of a cDNA microarray to determine molecular mechanisms involved in grey platelet syndrome. Br J Haematol. 2003; 122:142-149.
96. Tefferi A, Spivak JL. Polycythemia vera: scientific advances and current practice. Semin Hematol. 2005;42:206-220.
97. Pellagatti A, Vetrie D, Langford CF, et al. Gene expression profiling in polycythemia vera using cDNA microarray technology. Cancer Res. 2003; 63:3940-3944.
98. Goerttler PS, Kreutz C, Donauer J, et al. Gene expression profiling in polycythaemia vera: overexpression of transcription factor NF-E2. Br J Haematol. 2005;129:138-150.
99. Kralovics R, Teo SS, Buser AS, et al. Altered gene expression in myeloproliferative disorders correlates with activation of signaling by the V617F mutation of Jak2. Blood. 2005;106:3374-3376.
100. Wang HY, Malek RL, Kwitek AE, et al. Assessing unmodified 70-mer oligonucleotide probe performance on glass-slide microarrays. Genome Biol. 2003;4:R5.
101. Sauer S, Lange BM, Gobom J, Nyarsik L, Seitz H, Lehrach H. Miniaturization in functional genomics and proteomics. Nat Rev Genet. 2005; 6:465-476.
102. Hochstrasser DF, Sanchez JC, Appel RD. Proteomics and its trends facing nature's complexity. Proteomics. 2002;2:807-812.

Chapter 16

Transcript Profiling of Human Platelets Using Microarray and Serial Analysis of Gene Expression (SAGE)

Dmitri V. Gnatenko, John J. Dunn, John Schwedes, and Wadie F. Bahou

Abstract

Platelets are anucleated cells that are generated from megakaryocytes via thrombopoiesis. They lack genomic DNA but have a pool of individual mRNA transcripts. Taken together, these mRNAs constitute a platelet transcriptome. Platelets have a unique and reproducible transcript profile, which includes ~1,600–3,000 individual transcripts. In this chapter, we will focus on platelet purification and on transcript profiling using an Affymetrix microarray platform and serial analysis of gene expression (SAGE). Platelet purification is described in detail. Large-scale platelet purification schema is designed to purify platelets from apheresis platelet bags (~3–5 $\times$ 10^{11} platelets/bag). Modification of this schema – small-scale platelet purification – is designed to isolate platelets from 20 ml of peripheral blood. This chapter provides detailed protocols for microarray and SAGE transcript profiling. We also discuss peculiarities of platelet purification, RNA isolation, and transcript profiling.

Key words: Platelet RNA, microarray, serial analysis of gene expression.

1. Introduction

Human blood platelets are generated from precursor bone marrow megakaryocytes. They play critical roles in normal hemostatic processes and pathologic conditions such as thrombosis, vascular remodeling, inflammation, and wound repair. Platelets are anucleate and lack nuclear DNA, although they retain megakaryocyte-derived mRNAs *(1)*. Platelets also contain rough endoplasmic reticulum and polyribosomes, and preserve the ability for protein biosynthesis from cytoplasmic mRNA *(2)*. The entire pool of individual platelet mRNAs constitutes a platelet transcriptome. Historically, platelet transcriptome has been regarded as invariant and gradually declining with cell senescence.

Peter Bugert (ed.), *DNA and RNA Profiling in Human Blood: Methods and Protocols, vol. 496*

DOI 10.1007/978-1-59745-553-4_16 Springerprotocols.com

Indeed, younger platelets contain larger amounts of mRNA with a greater intrinsic capacity for protein biosynthesis, as determined using fluorescent nucleic acid dyes such as thiazole orange *(3)*. This traditional platelet/mRNA paradigm was challenged recently when a signal-dependent pre-mRNA splicing was identified in platelets *(4)*. Signal-dependent splicing provides a mechanism for altering the repertoire of translatable mRNAs in response to cellular activation/stimulation. Furthermore, platelets possess essential components of a functional spliceosome and selected unspliced pre-mRNAs. Platelet spliceosomes have a unique ability to splice pre-mRNA in the cytoplasm (as opposed to the typical nuclear location) – a capability not described in any other mammalian cell.

Modern high-throughput technologies such as microarray and serial analysis of gene expression (SAGE) allow sensitive and reliable transcript profiling and have been commonly used for such studies with a wide variety of cell types. To date, however, only a few studies have focused on similar analysis of the human platelet transcriptome. One difficulty in working with platelets is that they contain far fewer transcripts than most nucleated cells, ranging from ~1,600 to 3,000 mRNAs *(5–9)*. In particular, platelets express much less transcripts than leukocytes *(10, 11)* which can easily contaminate the platelet fraction derived from whole blood. To date, three important conclusions can be made from platelet transcript profiling. First, platelets have much fewer individual mRNA transcripts than most nucleated cells, presumably due to the lack of ongoing transcription. Second, platelets have a unique and reproducible transcript profile, distinctive from transcript profiles of any other cell types. Third, this profile is different for normal platelets and for platelets of patients with platelet-associated diseases, in particular of patients with essential thrombocythemia.

Taken together, these studies demonstrate that platelet transcript profiling can be applied to dissect molecular mechanisms of platelet function and to better understand and possibly aid in diagnosing certain platelet-associated diseases. In this chapter we will discuss protocols for platelet purification and transcript profiling, paying specific attention to the critical details and potential limitations of these protocols.

2. Materials

2.1. Platelet Purification

For preparation of all buffers and solutions, deionized and sterilized water (ddH_2O) should be used unless otherwise specified. Reagents and solutions are stored according to manufacturer's instructions. For all procedures involved in RNA isolation, SAGE,

and microarray analysis, ddH_2O should be treated with diethyl pyrocarbonate (DEPC), and all glass- and plastic-ware should be DEPC-treated to minimize RNA losses. Use pipette tips with aerosol filters, wear gloves and if possible perform all manipulations with RNA in a "RNAse-free" hood.

2.1.1. Platelet Purification from Platelet Apheresis Bags

1. Sodium ethylenediaminetetraacetate (EDTA) stock solution, 0.5 M, pH 8.0.
2. Sodium citrate, 4% (w/v).
3. Phosphate-buffered saline (PBS), pH 7.4: 137 mM NaCl, 2.7 mM KCl, 4.3 mM Na_2PO_4, 1.4 mM KH_2PO_4.
4. HEPES-buffered modified Tyrodes buffer (HBMT): 10 mM HEPES pH 7.4; 150 mM NaCl; 2.5 mM KCl; 0.3 mM NaH_2PO_4; 12 mM $NaHCO_3$; 0.2% BSA; 0.1% glucose; 2 mM EDTA.
5. Triton X-100.
6. Prostaglandin E (PGE) (Sigma, St Louis, MO, USA): 1 mg/ml stock solution in 95% ethanol, 2.8 mM). Solution is light-sensitive, keep the tube wrapped in foil at –20°C.
7. Aspirin (Sigma).
8. Aqueous chromatography column (5.0 cm × 60 cm; Spectrum Chromatography, Houston, TX, USA).
9. BioGel A50 M (BioRad) or Sepharose CL-4B (Amersham).
10. Mouse anti-CD45 antibody, conjugated to magnetic beads (Miltenyi Biotec Inc., Auburn, CA, USA).
11. MidiMacs starting kit (Miltenyi Biotec) which includes MidiMacs magnetic columns (Miltenyi Biotec) and a magnetic stand.
12. BlueMax™ polypropylene conical tubes, 50 ml (Falcon; BD, Franklin Lakes, NJ USA).
13. Pore size leukocyte membrane filters, 5 μm.

2.1.2. Characterization of the Efficiency of Platelet Purification

1. FITC-conjugated anti-CD41b (BD Pharmingen).
2. PE-conjugated anti-glycophorin (BD Pharmingen).
3. PERCP-conjugated anti-CD45 (BD Pharmingen).
4. Formalin solution (1%) in PBS.
5. Leukocyte contamination of platelet preparation can also be addressed at RNA level by reverse transcription–polymerase chain reaction (RT–PCR) using primers specific to leukocyte common antigen CD45 (NCBI accession number Y00638):
   Forward, 5′-GCTCAGAATGGACAAGTA-3′
   Reverse, 5′-CACACCCATACACACATACA-3′

2.1.3. Platelet Purification from 20 ml of Peripheral Blood

1. Prepare in advance: 20-ml disposable syringe; 19-gauge needle; two columns for platelet gel-filtration (commercially available 60-ml disposable syringes is a simple, inexpensive, and durable columns for platelet gel filtration).
2. Place a circular piece of nylon mesh with 85 micron pore size (Small Parts, Inc.) at the bottom.

3. Place the rubber ring from the plunger of the syringe on top of the mesh to hold it in place.
4. Load suspension of Sepharose 2B beads (Amersham) in HBMT buffer until bead volume reaches 50 ml.
5. Equilibrate the column by passing through 150 ml of HBMT buffer, pH 7.45.
6. Seal top and bottom of the column and keep it at room temperature until use.

2.2. Platelet RNA Isolation

1. Trizol® reagent, (Invitrogen).
2. Chloroform.
3. Isopropanol.
4. Ethanol (100%).
5. "Glyco-blue" (Ambion).
6. SuperRNAseIn™ RNAse inhibitor, 20 U/μl (Ambion).
7. DEPC (Sigma). Treat 1.5- and 0.5-ml Eppendorf tubes with DEPC, autoclave, and store them in a closed beaker.
8. For RNA isolation from platelets isolated from an apheresis bag, treat four 30-ml Corex glass centrifuge tubes with DEPC, autoclave, and store in aluminum foil until use. Use glass pipettes to measure organic solvents.
9. For RNA isolation, use only DEPC-treated (or RNAse-free) plastic- and glass-ware. Prepare DEPC-treated H_2O to dissolve RNA pellets.

2.3. Microarray Profiling

For preparation of all buffers and solutions, deionized and sterilized DEPC-treated water (ddH_2O) should be used unless specified otherwise. Store reagents and solutions according to manufacturer's instructions or as described in Current Protocols in Molecular Biology.

1. $T7dT_{24}$ primer (Proligo).
2. SuperScript II reverse transcriptase (Invitrogen).
3. DEPC-treated H_2O.
4. dNTP mix (10 mM) prepared from individual 100 mM dATP, dCTP, dGTP, dTTP (Invitrogen).
5. *E.coli* DNA ligase (10 U/μl; Invitrogen).
6. DNA polymerase I (10 U/μl; Invitrogen).
7. RNAse H (1–4 U/μl; Invitrogen).
8. T4 DNA polymerase (5 U/μl; Invitrogen).
9. EDTA (0.5 M) in DEPC-treated H_2O.
10. Phaselock tube (Eppendorf).
11. Phenol, chloroform, NH_4O-acetate, absolute ethanol.
12. BioArray High Yield kit (Enzo Biochem).
13. RNeasy mini kit (Qiagen).
14. Dnase I solution (RNase-free DNase set; Qiagen).
15. Fragmentation buffer: 200 mM Tris–acetate, pH 8.1; 500 mM KOAc; 150 mM MgOAc.
16. Control oligonucleotide B2, 20× eukaryotic hybridization controls, 2× MES hybridization buffer all from Affymetrix.

17. BSA (50 mg/ml; Invitrogen).
18. Herring sperm DNA (10 μg/μl; Promega).

2.4. Serial Analysis of Gene Expression

All buffers should be prepared using DEPC-treated ddH_2O.

1. Glycogen, 5 μg/μl (Invitrogen).
2. Phenol:chloroform:isoamyl alcohol.
3. SuperRNAse-In, 20 U/μl (Invitrogen).
4. GlycoBlue (Ambion).
5. Spermidine $(HCl)_3$ (Sigma).
6. *E.coli* DNA ligase, *E.coli* DNA polymerase, *E.coli* RNAse H (SuperScript Choice System for cDNA synthesis; Invitrogen).
7. *E.coli* exonuclease I, 20 U/μl (NEB).
8. TEN/BSA buffer: 10 mM Tris–HCl, pH 8.0; 1 mM EDTA; 1 M NaCl; 1% BSA.
9. *Mme*I restriction buffer (10×): 100 mM HEPES, pH 8.0; 25 mM K-acetate, pH 8.0; 50 mM Mg-acetate, pH 8.0; 20 mM DTT.
10. *Mme*I restriction enzyme (NEB).
11. *Nla*III restriction enzyme (NEB) (Note: this enzyme should be aliquoted and then stored at –80°C).
12. SAM (100×): 4 mM SAM (*S*-adenosylomethionine hydrochloride).
13. Ammonium acetate (7.5 M).
14. *Prepare in advance:* Oligo$(dT)_{25}$-conjugated magnetic beads (Dynal).
15. Dynabeads mRNA DIRECT kit (includes Dynabeads Oligo$(dT)_{25}$ magnetic beads, lysis/binding buffer, washing buffers A and B; Dynal).
16. Magnetic Particle Concentrator MPC-S (Invitrogen or Dynal).
17. Primers and cassettes:
    cassette A primers:
    5′-TTTGGATTTGCTGGTCGAGTACAACTAGGCTTAAT CCGACATG-3′
    5′-pGTCGGATTAAGCCTAGTTGTACTCGACCAGCAAA TCC-3′amino modified
    cassette B primers:
    5′-pTTCATGGCGGAGACGTCCGCCACTAGTGTCGCA ACTGACTA-3′
    5′-TAGTCAGTTGCGACACTAGTGGCGGACGTCTCCG CCATGAANN-3′
18. Primers for PCR amplification:
    Forward, 5′- biotinylated, corresponding to a portion of cassette A top strand: 5′-Biotin-GGATTTGCTGGTCGAG TACA-3′
    Reverse, non-biotinylated, corresponds to a portion of cassette B bottom strand: 5′-TAGTCAGTTGCGACACT AGTGGC-3′

3. Methods

3.1. Platelet Purification

Initially, we have developed a large-scale platelet purification schema to isolate platelets starting from an apheresis bag (~3–5 × 10^{11} platelets) *(6)*. This schema includes four consecutive steps, each based on a different principle of cell separation. Combining the different separation techniques in one purification scheme allows for highly efficient platelet purification (*see* **Note 1**). The resulting product should contain only 5–10 leukocytes per 1 × 10^5 platelets. These steps include *(1)* centrifugation; *(2)* gel-filtration; *(3)* filtration; and *(4)* $CD45^+$-immunodepletion. Combined, these steps complement each other to achieve the highest purity of platelets while minimizing platelet losses. Large-scale schema allows isolation of ~ 50–100 μg of total RNA from highly purified platelets. This amount is sufficient for direct high-throughput transcript profiling by microarray and SAGE technology. The major drawbacks of this schema are that: (i) it is time- and labor-consuming; (ii) it cannot be used for screening large groups of patients, and (iii) it requires researcher access to properly stored fresh apheresis bags and to sophisticated laboratory equipment.

To adapt platelet purification for processing large number of samples, the schema may be modified to isolate platelets from 20 ml of peripheral blood. The modified procedure is simple, reliable and can be used for screening a large number of patients, and yet this schema is efficient in removing contaminating cells. Its main drawback is that it does not provide enough total RNA for direct transcript profiling (**Fig. 16.1**). Platelet purification from 20 ml of peripheral blood using the platelet-rich plasma followed by gel-filtration yields ~2 × 10^9 platelets total. Yield of total platelet RNA varies between 1.0 and 2.5 μg, whereas accurate and reproducible transcript profiling using Affymetrix technology requires ~ 5 μg of total RNA. Although this amount is insufficient for microarray or SAGE analyses, it can be used for the analysis of expression of selected transcript using quantitative real-time RT-PCR (Q-PCR) technique.

Alternatively, total platelet RNA can be converted to cDNA, amplified by PCR or in vitro transcription and then used for transcript profiling. Although amplification can introduce some variance in original transcript ratios, it has been successfully applied for transcript profiling (reviewed at Lockhart et al. *(12)*). In our experience, in vitro amplification is efficient and reproducible. Starting with 60 ng of total platelet RNA and using Ovation Aminoallyl RNA amplification and labeling system (NuGENE), 4–5 μg of labeled cDNA can usually be generated. This amount is sufficient for transcript profiling using custom-made

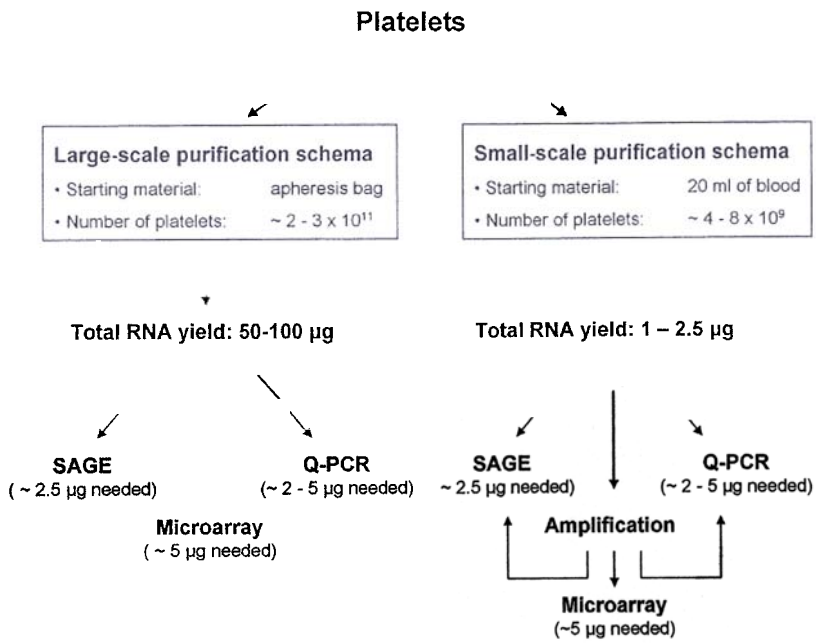


Fig. 16. Overview of platelet purification schema.

oligonucleotide platelet-specific microarray chips. Moreover, hybridization of amplified and non-amplified platelet RNA to this chip reveals excellent concordance (**Fig. 16.2**). Potentially, this cDNA can also be hybridized to standard Affymetrix chips.

3.1.1. Platelet Purification from Platelet Apheresis Bags

The volume of liquid and platelet content of each apheresis bag varies depending on many factors, including donor, techniques used for apheresis, and other variables. To standardize the purification schema, the same starting volume (150 ml) should be used

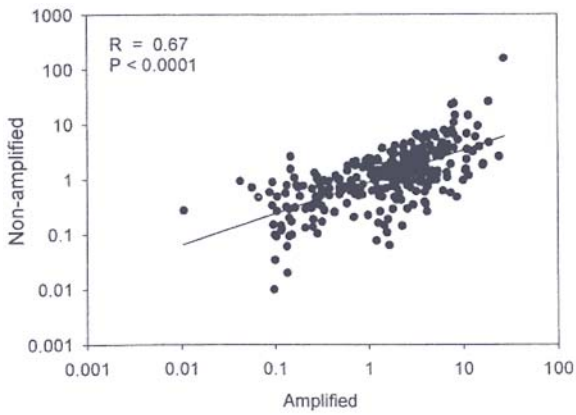


Fig. 16.2. Correlation between amplified and non-amplified samples of total platelet RNA, analyzed by hybridization to a customized platelet-specific microarray chip (R=0.67, P<0.0001).

for every purification. This is an optimal volume for efficient gel-filtration on the column of given size (5 × 60 cm, 1,000 ml support). The remainder of the platelets from the apheresis bag can either be discarded or used for other studies.

For platelet isolation from an apheresis bag, prepare a column for platelet gel-filtration in advance:

1. For large-scale platelet purification a large aqueous chromatography column is packed with BioGel A50M or Sepharose CL-4B in accordance with manufacturer's instructions.
2. Remove air bubbles from the bead slurry by applying a vacuum for 5–7 min before loading into the column. Gel-filtration medium of 1,000 ml bed is used to purify platelets from one apheresis bag.
3. Wash the settled bed in the column by passing through 500 ml of HBMT buffer, containing 5 mM EDTA, 0.05 mM aspirin, and 0.1 μM PGE.
4. Assemble a chromatography system by connecting a pump proximal to the column and an optical unit to read eluate absorbance at 280 nm. Make sure the system is hermetically closed and connected to a pump providing a liquid flow rate from 60 to 600 ml/h. If possible, use a fraction collector to harvest the eluate and a suitable UV recorder to monitor and record the platelet's (A_{280}) elution peak.

   Start with one fresh platelet apheresis bag:
5. Take 150 ml of platelet concentrate (typically 2–3 × 10^{11} platelets), place it into a plastic container and add the following inhibitors of platelet activation: 1.5 ml of 0.5 M EDTA (final concentration = 5 mM); 600 μl of 12.5 mM aspirin (final concentration = 0.05 mM); 7 μl of PGE (final concentration = 0.1 μM), mix well by gentle stirring. These additions are very important in order to prevent platelet aggregation throughout the entire procedure since aggregation complicates platelet purification and dramatically reduces the amount and quality of the resulting platelet RNA.
6. At this point, take a 10 μl sample (Sample 1) for platelet count and flow cytometry analysis.
7. Divide the 150 ml of platelet concentrate equally among three 50-ml falcon tubes and centrifuge at 140*g* for 15 min at room temperature using a swinging bucket rotor centrifuge.
8. Once centrifugation is complete, gently take upper 90% (45 ml) of platelet concentrate, leaving ~ 3–5 ml at the bottom of each tube. This centrifugation efficiently removes a large fraction of contaminating leukocytes.
9. Combine platelet concentrate from three tubes in another plastic container, mix well and take a 10 μl sample (Sample 2) for the analysis. This procedure takes about 20–25 min.

   Start with gel-filtration:

10. Open the chromatography column and remove upper part of HBMT buffer, leaving just enough liquid to cover the beads.
11. Let remaining buffer enter the gel by gravity force and load platelet concentrate (~140 ml) without allowing the bed to dry out.
12. Let platelet concentrate completely enter the gel. If necessary, use the pump.
13. Immediately after the platelets enter the gel, wash column walls with 10 ml of HBMT buffer and let it enter the gel.
14. Then add 50–70 ml of buffer to the top of the gel, close the column and turn on the pump and optical unit to measure the platelet elution profile.
15. Maintain a liquid flow rate at 400–600 ml/h. Monitor movement of platelets through the gel and the liquid flow rate. If platelets become activated, they will clot the column, i.e., liquid flow decreases and buffer starts to accumulate above the gel. If the procedure goes normally, the quiescent platelets will elute from the column as a single wide peak.
16. Stop collecting platelets when the optical density reaches approximately one-third of its maximal height to avoid contamination with plasma proteins, which elute somewhat later than the platelet peak. The volume of the pooled platelet fractions after the column varies between 150 and 200 ml.
17. Combine all the gel filtered platelets in one plastic container, add aspirin to final concentration 0.05 mM and PGE to 0.1 μM, mix well and take another 10 μl sample (Sample 3) for analysis. It takes between 90 and 120 min to complete chromatography from the moment the platelet concentrate is loaded onto the gel. This procedure transfers platelets from plasma into HBMT buffer, removes cell aggregates and further reduces leukocyte contamination.

    Proceed with leukocyte filtration:
18. Assemble a filtration unit by placing a 5 μm pore size leukocyte membrane filter and securing it in its holder.
19. Pre-wet the membrane with HBMT buffer, but do not apply vacuum.
20. Load platelet concentrate and let it pass through the membrane by gravity flow. This procedure takes from 5 to 10 min, the volume of platelet concentrate does not change, and platelet losses are minimal.
21. Once platelet concentrate passes through the membrane, take a 10 μl aliquot (Sample 4) for analysis. This step is designed to further reduce the amount of contaminating leukocytes in the platelet preparation.

    Proceed with $CD45^+$cell immunodepletion (*see* **Note 2**):
22. Place platelet concentrate in corresponding number of 50 ml falcon tubes.

23. Harvest platelets by centrifugation using a swinging bucket rotor at 1,500*g* for 15 min at room temperature.
24. Discard supernatant and re-suspend platelets gently but thoroughly by adding 10 ml of HBMT buffer to the first tube and serially transferring it to the next tube until all platelets are re-suspended.
25. Gently pipette platelet concentrate to break any visible clumps or aggregates.
26. Bring the total volume to 20 ml and add 120 µl of mouse anti-CD45 antibody conjugated to magnetic beads.
27. Incubate at room temperature with mixing on a low-speed rotating platform for 45 min.
28. While platelets are incubating, prepare two magnetic columns. Place them on magnets attached to magnetic stands, load 5 ml of HBMT buffer and let the buffer pass through columns by gravity. Do not use a plunger to push buffer through the columns.
29. Once incubation of the platelet fraction with antibody is complete, apply 3–4 ml of the mixture onto the column and let it enter the magnetic mesh. Continue adding platelet concentrate as purified platelets elute from the bottom of the columns.
30. Use plungers and 5 ml of HBMT buffer to displace the entire sample.
31. Combine platelet and HBMT wash fractions and take 10 µl of the material (Sample 5) for analysis.
32. Collect platelets by centrifugation at 1,500*g* for 15 min at room temperature in 50-ml falcon tube. Usually the volume of a platelet pellet is about 5 ml. At this point, the platelet pellet can be frozen at –80°C or used immediately for total RNA isolation.

3.1.2. Characterization of the Efficiency of Platelet Purification

The yield and efficiency of platelet purification should be documented at each step. To analyze platelet yield, obtain platelet count for samples 1–5 and calculate platelet number at each step of purification. Typical, overall platelet yield is ~ 80%. Purification efficiency can be analyzed by flow cytometry.

1. Incubate aliquots containing 2×10^6 platelets from sample 1–5 with saturating concentrations of FITC-conjugated anti-CD41, PE-conjugated anti-glycophorin, and PERCP-conjugated anti-CD45 for 15 min in the dark at 25°C.
2. Wash them once with PBS, and fix in 1× PBS/1% formalin.
3. Analyze samples using a flow cytometer; quantify the number of CD45- and glycophorin-positive events in each sample.
4. Characterize efficiency of platelet purification by expressing it as the number of CD45-positive events (leukocytes) per

100,000 CD41-positive events (platelets). In our experience, platelet purity varied from 3 to 10 leukocytes per 1×10^5 platelets.

Leukocyte contamination of platelet preparation can also be addressed by RT-PCR using a variety of primers specific to leukocyte-expressed transcripts. Primers specific to leukocyte common antigen CD45 (NCBI accession number Y00638) have been successfully used to evaluate efficiency of leukocyte removal *(6, 9)*.

3.1.3. Platelet Purification from 20 ml of Peripheral Blood

1. Prepare in advance: 20-ml disposable syringe; 19-gauge needle; two columns for platelet gel-filtration (commercially available 60-ml disposable syringes is a simple, inexpensive and durable columns for platelet gel filtration).
2. Place a circular piece of nylon mesh with 85 micron pore size at the bottom.
3. Place the rubber ring from the plunger of the syringe on top of the mesh to hold it in place.
4. Load suspension of Sepharose 2B beads in HBMT buffer until bead volume reaches 50 ml.
5. Equilibrate the column by passing through 150 ml of HBMT buffer, pH 7.45.
6. Seal top and bottom of the column and keep it at room temperature until use.
7. Collect blood sample by venipuncture using 19-gauge needle attached to 20-ml syringe.
8. Divide it in two 10-ml aliquots, place each into a plastic container and add 1 ml of 100 mM EDTA to each aliquot as an anti-coagulant.
9. Place 1 µl of blood into a small Eppendorf tube containing 20 µl of 500 mM EDTA for complete blood count (CBC).
10. Generate platelet-rich plasma (PRP) by centrifugation of blood sample for 3.5 min at 700*g* at room temperature.
11. Gently take upper 9/10 of PRP (*see* **Note 3**). The total volume of PRP from 20 ml of blood typically varies between 4 and 8 ml.
12. Gently apply 2–3 ml PRP to each of two Sepharose 2B columns and allow platelets to enter the columns by gravity flow (*see* **Note 4**).
13. Once platelets enter, fill the columns with HBMT buffer to the top. Continue adding buffer during gel filtration to avoid drying of the beads.
14. As platelets elute they have a visible cloudiness that can be detected by eye; pool these fractions. The combined volume from two columns should be from 6 to 11 ml.
15. Harvest the platelets by centrifugation for 10 min at 2,000*g*.
16. Remove as much supernatant as possible, being careful not to disturb the soft platelet pellet.

17. At this point platelets can either be stored at −80°C or immediately used for RNA isolation. With reference to RNA instability, immediate RNA isolation may be a better solution.

3.2. Platelet RNA Isolation

3.2.1. Total RNA Isolation from Apheresis Platelet Bag

1. Thaw platelet pellet from the previous purification step by incubating falcon tube in room temperature water for 3–5 min.
2. Add 10 ml of Trizol reagent using a plastic pipette. If platelet pellet was not frozen, add 10 ml of Trizol directly to the pellet.
3. Use the plastic pipette to break up any clumps which tend to accumulate at the bottom of the tube in the dissolved pellet. It may take several minutes to completely disintegrate large clumps.
4. Add another 10 ml of Trizol to falcon tube, close it tightly and shake well by hand.
5. Transfer lysate into two 30-ml DEPC-treated Corex tubes, aliquoting equal volumes of platelet lysate per tube.
6. To each Corex tube, add another 10 ml of Trizol reagent. Finally, platelet pellet from an apheresis bag should be dissolved in 40 ml of Trizol, divided between two 30 ml Corex tubes. In our experience, we have found that for a 5-ml platelet pellet, the optimal amount of Trizol to be added is 40 ml (*see* **Note 5**).
7. Incubate tubes at room temperature for 5 min.
8. To each tube, add 4 ml of chloroform using a glass pipette. Mix well and incubate for 3 min at room temperature. (Note: parafilm dissolves in chloroform. If necessary seal tops with Dura Seal Laboratory Stretch film from Diversified Biotech, Boston.).
9. Centrifuge at 12,000*g*, 4°C for 15 min.
10. Gently take upper nine-tenths of clear supernatant layer and transfer to two new DEPC-treated Corex tubes; avoid touching protein interphase.
11. To each tube, add 2 µl of GlycoBlue as an indicator of RNA precipitation. Mix well.
12. Add 10 ml of cold (0°C) isopropanol. Close tube tightly with parafilm or Dura Seal, shake well by hand. At this point, tubes can be stored overnight at −20°C.
13. Precipitate total RNA by centrifugation at 12,000*g* at 4°C for 30 min.
14. Gently remove and discard supernatant.
15. Wash platelet RNA pellet – add 5 ml of 75% ethanol per Corex tube, wash tube walls using a pipette, and centrifuge at 12,000*g* for 15 min at 4°C.
16. Repeat washing step one more time.

17. After final wash, remove supernatant and air-dry RNA pellet for 3–5 min. Do not let RNA pellet dry completely since it may make re-suspension of RNA difficult.
18. Carefully re-suspend RNA from both Corex tubes in 100 µl of DEPC-treated water.
19. Transfer RNA solution into DEPC-treated microcentrifuge tube, add 1 µl of RNAse inhibitor – RNAseIn and mix well.
20. Wash walls of Corex tubes with another 50–100 µl of DEPC-treated water, combine with RNA solution, and mix well.
21. Incubate RNA solution at 55°C for 5 min to aid solubilization, quickly transfer onto ice.
22. Gently pipette RNA solution and take 5 µl for quantity and quality analyses. At this point, RNA can be aliquoted in smaller portions and stored at –80°C for further analysis.
23. In our experience, 50–100 µg of total RNA can typically be isolated from 150 ml of an apheresis bag. Total RNA yield for patients with essential thrombocythemia can reach up to 260 µg. This amount is more than sufficient for reliable transcript profiling using both microarray and SAGE.
24. To quantify RNA yield, use a suitable spectrophotometer – we use a Biophotometer from Eppendorf. To address quality of RNA, use a BioAnalyzer (Bio-Rad). A typical capillary electrophoresis of total platelet RNA is shown at **Fig. 16.3**. Note similarities between total RNA isolated from human liver and from platelets.
25. Once purified and characterized, platelet RNA with added RNAse inhibitors can be stored at –80°C for several months without detectable degradation.

3.2.2. Total RNA Isolation from Peripheral Blood Samples

1. Add 600 µl of Trizol to the platelet pellet, thoroughly re-suspend platelets and incubate for 3 min. Follow Trizol protocol after this step.
2. Add 120 µl of chloroform, shake tubes vigorously by hand for 15 s and incubate at room temperature for 3 min.
3. Centrifuge the sample at 12,000*g* for 15 min at 2–8°C. Following centrifugation, the mixture separates into a lower red, phenol–chloroform phase, an interphase, and a colorless upper aqueous phase; RNA remains in aqueous phase.
4. Take upper phase, avoiding cross-contamination with interphase material and transfer into fresh DEPC-treated 1.5 ml microcentrifuge tube. At this point the total volume is typically ~700 µl.
5. Add 1 µl of Glyco-blue to the aqueous layer to facilitate RNA precipitation. Mix sample well.
6. To precipitate RNA, add 600 µl of ice-cold isopropyl alcohol, close tube tightly, shake it well by hand and briefly vortex. Incubate sample for 30 min (or overnight) at –20°C.
7. Mix well and centrifuge at 12,000*g* for 15 min at 2–8°C.

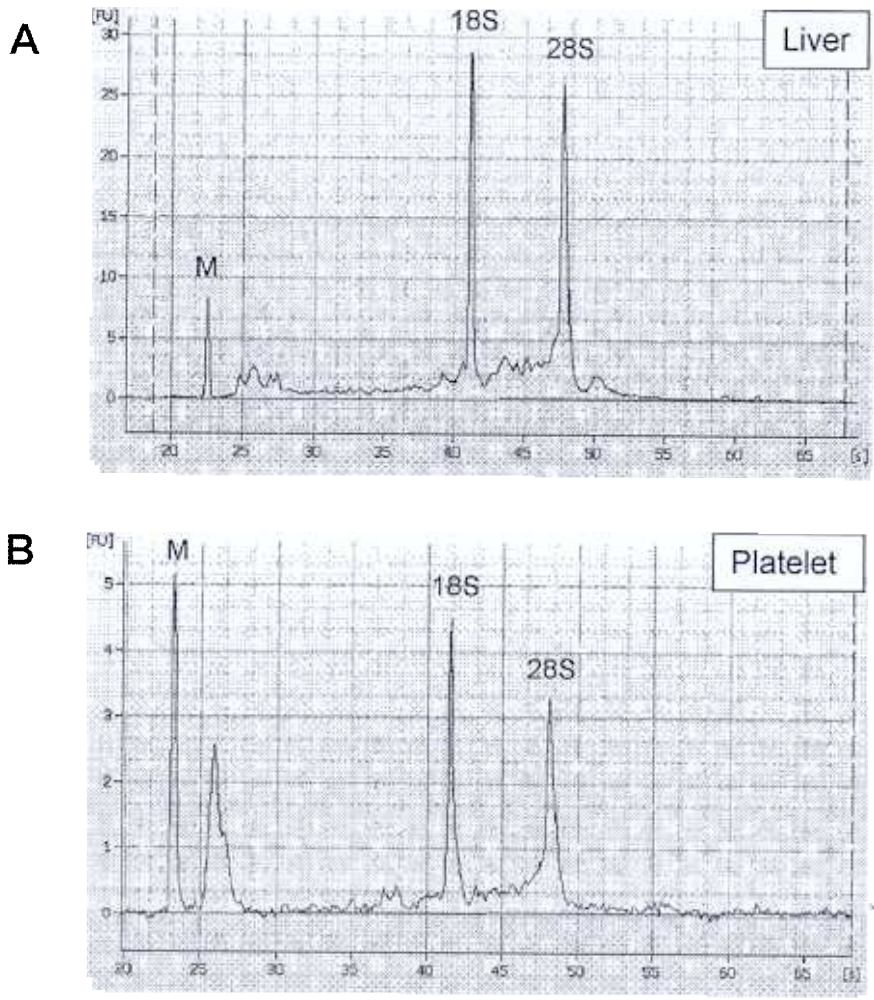


Fig. 16.3. Analysis of total RNA by capillary electrophoresis using BioAnalyzer Model 2,100. Total RNA from liver and from platelets were analyzed under the similar conditions in two separate experiments. (**A**) Adult human male total RNA (Stratagene); (**B**) Total platelet RNA. M – Agilent RNA 6,000 nanomarker. Ribosomal RNA (18S and 28S) peaks are present in both samples.

8. Carefully remove supernatant – at this point RNA pellet will be visible only due to the blue dot formed by Glyco-blue co-precipitation. Be careful not to discard RNA pellet.
9. Wash RNA pellet twice with 1 ml of 75% ice-cold ethanol: add 75% ethanol, vortex briefly, and centrifuge for 5 min at 12,000*g* at 2–8°C. Gently remove the supernatant and repeat the 75% ethanol washing. Exercise extra caution not to discard RNA pellet which at this point is almost invisible. (We routinely save the washes in case of RNA loss).
10. Once washing is complete, gently remove any residual liquid and air-dry pellets for 10–15 min at room temperature.

11. To re-suspend the RNA pellet, add 50 µl of DEPC-treated H_2O containing 1 µl of SuperRNAseIn to inhibit RNAses. Slowly pipette RNA solution up and down using a sterile pipette tip equipped with aerosol filter.
12. Incubate RNA at 53°C for 5 min to aid resolubilization, quickly transfer on ice, incubate for 5 min and briefly centrifuge at 5,000*g* to collect any vapor condensate from the walls of the tube.

3.3. Platelet Transcript Profiling Using Affymetrix Microarray Platform

Development of global mRNA profiling technologies *(13, 14)* (such as microarray and SAGE), coupled with recent completion of the Human Genome project provides novel approaches to study platelet function *(15, 16)*. Microarray transcript profiling uses grids of "probes", specific to known DNA sequences, to capture and quantify RNA transcripts of corresponding genes. To date, many different microarray platforms have been developed, all of them include (1) isolation of RNA from biological sample; (2) copying of RNA into either cRNA or cDNA, simultaneously incorporating either fluorescent nucleotides or a tag that is later used for fluorescent labeling; (3) hybridization to a microchip containing a grid of DNA probes; (4) washing; (5) labeling (if necessary, depending on the protocol); (6) scanning under laser light; and (7) image processing, data extraction, and analysis *(17)*. Hybridization to microarrays and quantitative real-time PCR are "closed-architecture" tools, i.e., these methods rely totally on pre-established sequence information whereas SAGE and similar techniques which do not require any a priori knowledge of the sample's structure or complexity are considered "open-architecture" tools. Each method has its own strengths and weaknesses and their utility is somewhat sample dependent. In our experience, hybridization to microarrays and quantitative real-time PCR are the methods of choice for analyzing platelet transcriptomes.

3.3.1. cDNA Synthesis

The cDNA synthesis is performed using reagents from Invitrogen Inc.

1. Mix 5 µg total RNA with 1 µl (100 pmol) $T7dT_{24}$ primer and DEPC-treated water to a final volume of 12 µl in 1.5 ml tube. Heat to 70°C for 10 min. Quick spin then place on ice.
2. Prepare a master mix with a 10% excess. For each sample mix 4 µl of 5× first strand buffer, 2 µl of 0.1 M DTT (first strand buffer and DTT supplied with Super Script II RT) and 1 µl of 10 mM dNTP mix (2.5 mM each dNTP).
3. Add 7 µl of the master mix to each sample tube of total RNA and primer. Mix and place in a 42°C bath for 2 min.
4. Add 1 µl (200U) of Super ScriptIIRT and incubate at 42°C for 1 h.

5. Prepare a master mix for the second strand synthesis containing: 91 µl DEPC-treated H_2O (Ambion), 30 µl second strand buffer (Invitrogen), 3 µl of 10 mM dNTP, 1 µl (10 U) DNA ligase, 4 µl (40 U) DNA polymerase I, and 1 µl (2 U) RNaseH. Cool a bath to 16°C.
6. Add 130 µl of the second strand master mix to each sample mix and spin and place into the 16°C bath for 2 h.
7. Add 2 µl (10 U) T4 DNA polymerase, incubate at 16°C for 5 min. Add 10 µl of 0.5 M EDTA to each sample. At this point the samples can be frozen at –20°C or purified.

3.3.2. cDNA Purification

1. Label a Phaselock tube for each sample and spin the tubes at 12,000*g* for 25 s. Add 162 µl of phenol/chloroform (21:1) to each sample and mix. Transfer the entire 324 µl of sample to a Phaselock tube and spin for 2 min at 12,000*g*.
2. Carefully transfer the supernatant to a new 1.5 ml tube. Add 0.5 volumes 7.5 M NH_4OAc and 2.5 volumes cold absolute ethanol. Mix and centrifuge for 20 min at > 14,000*g*.
3. Remove the supernatant – a faint, small white pellet should be visible. Wash the pellet with 1 ml of 80% ethanol and centrifuge for 15 min at > 14,000*g*.
4. Remove all of the supernatant and air-dry the pellets for a few minutes.
5. Re-suspend the pellet in 22 µl DEPC-treated H_2O.

3.3.3. IVT Reaction

1. The BioArray High Yield kit is used to generate the cRNA.
2. Mix 18 µl of the master mix and the 22 µl of sample, aliquot into several tubes, and incubate at 37°C overnight.

3.3.4. cRNA Purification

The cRNA is purified using the RNeasy mini kit (Qiagen).

1. Add 60 µl DEPC-treated water and 350 µl buffer RLT (do not add the BME to the buffer RLT) mix.
2. Add 250 µl of 100% EtOH, mix by pipetting several times and transfer to a Qiagen RNeasy column.
3. Centrifuge 18 s at 12,000*g*. Reapply the sample and spin again for 18 s at 12,000*g*.
4. Discard tube and flow-through. Place column in a new tube. Apply 350 µl buffer RW1 to the column. Spin 18 s at 12,000*g*. Discard flow-through.
5. Carefully apply 80 µl DNAse I solution (10 µl DNAse I + 70 µl RDD buffer. Incubate 15 min at room temperature.
6. Apply 350 µl buffer RW1. Centrifuge at 12,000*g* for 18 s.
7. Discard tube and flow-through. Place column in a new tube and apply 500 µl buffer RPE. Centrifuge at 12,000*g* for 18 s. Discard tube and flow-through.
8. Place column in a new tube and apply 500 µl buffer RPE and centrifuge 2 min at 14,000*g*. Discard tube and flow-through

making sure that there is no residual flow-through on the column.

9. Place column in a labeled new microfuge tube. Carefully apply 50 µl DEPC-treated H_2O to the center of the column. Centrifuge 1 min at 12,000*g*. Carefully apply 50 µl DEPC-treated H_2O to the center of the column a second time. Centrifuge 1 min at 12,000*g*. Use of H_2O at 37°C improves cRNA recovery.
10. Add 10 µl of 5 M NH_4Oac and 275 µl of –20°C absolute EtOH. Incubate at –20°C for 30–60 min. Centrifuge at 18,000*g* at 4°C for 25 min. A large white pellet will be visible. Discard supernatant.
11. Wash pellet with ice-cold 70% EtOH, centrifuge at 18,000*g* for 15 min. Remove all supernatant and air-dry pellet.
12. Once all the EtOH as evaporated re-suspend the pellets in 32 µl DEPC-treated H_2O.
13. Measure nucleic acid concentration on a spectrophotometer.

3.3.5. Fragmentation of cRNA

1. In a microfuge tube combine 20 µg cRNA and DEPC-treated H_2O to a final volume of 32 µl.
2. Add 8 µl fragmentation buffer.
3. Incubate at 94°C for 35 min. Be sure to secure the top of the tube so it does not open during the incubation.
4. Place tubes on ice.

3.3.6. Hybridization of Arrays

1. Create a master mix of the hybridization cocktail containing: 15 ug fragmented cRNA, 5 µl control oligonucleotide B2, 15 µl 20× eukaryotic hybridization controls, 3 µl BSA 50 mg/ml, 3 µl sonicated herring sperm DNA 10 mg/ml, 150 µl 2× MES hybridization buffer (prepared as per Affymetrix GeneChip Expression Analysis Technical Manual.).
2. Add DEPC H_2O to a final volume of 300 µl.
3. Pre-hybridize the arrays with 1× MES hybridization buffer (Affymetrix GeneChip Expression Analysis Technical Manual.) at 42°C at 60 rpm in the Affymetrix Hybridization Oven Model 420.
4. Heat hybridization cocktail to 98°C for 5 min.
5. Transfer tubes to 42°C for 5 min.
6. Centrifuge samples at 18,000*g* for 5 min.
7. Remove arrays from hybridization oven and remove the 1× MES hybridization buffer.
8. Load the appropriate volume hybridization cocktail – 80 µl for Affymetrix Test 3 arrays chip or 200 µl for Affymetrix GeneChip. In our studies we used HGU133A microarray chip. Currently it is discontinued and replaced with more advanced microarray chips, which allow better quantification

of gene expression and include probes for more transcripts. For details of hybridization to these chips, refer to the manufacturer's instructions.

9. Place a "tough spot" over the lower loading port as this port leaks sometimes.
10. Place arrays into carriers and into the Affymetrix Hybridization Oven Model 420 set to 42°C at 60 rpm. Hybridize the arrays overnight ~16 h.

3.3.7. Washing and Staining the Arrays

1. Prepare wash buffers as per instructions in the Affymetrix GeneChip Expression Analysis Technical Manual.
2. Prepare 2× Stain 1 and 1× Stain 2 as per Affymetrix GeneChip Expression Analysis Technical Manual.
3. Remove the arrays from the Affymetrix Hybridization Oven, remove the tough spot.
4. Remove the hybridization cocktail and store at –80°C.
5. Load to appropriate volume of "buffer A" – 100 µl for the Test 3 array or 250 µl for the U133A array.
6. Wash and stain the arrays using the Micro v1.1 protocol for the Test 3 array and the EukGE-WS2v4 for the GeneChip U133.
7. When the third wash cycle begins start the scanner so that it completes its warm up cycle.

3.3.8. Scanning and Image Analysis

1. After the washing and staining protocol is complete check the arrays to make sure there are no air bubbles present.
2. Scan the arrays. Once the scan is complete check the grid alignment in the corners and in the center of the array.
3. Analyze the image using Microarray Analysis Suite 5.1 (Affymetrix) utilizing the following parameters scaling: All Probe sets: Target Signal 250, Normalization: User Defined, value 1. All other parameters are set at default values. The percent of genes called present (%*P*) is usually low in platelets (<20%), whereas the 5'/3' ratios is higher. The latter may be the result of partial degradation of platelet mRNA due to platelet aging and/or RNAse activity in the absence of ongoing transcription.
4. Further data processing and bioinformatic analysis is performed using GeneSpring software or its analogs using standard microarray data mining algorithms. Description of bioinformatic techniques is beyond the scope of this chapter.

3.4. Serial Analysis of Gene Expression Analysis

SAGE is a powerful mRNA-based method for comprehensive analysis of gene expression patterns *(18,19)*. In the original SAGE procedure, double-stranded cDNA is synthesized from poly $(A)^+$ mRNA by priming first-strand cDNA synthesis with a biotinylated oligo $(dT)_{18}$ primer. The cDNA is then cut with a restriction endonuclease having a 4-bp recognition sequence (typically *Nla*III, recognition sequence CATG, which theoretically results

in cleavage on average every 256 bp), and the 3'-terminal cDNA fragments are captured on streptavidin-coated magnetic beads. These fragments are ligated with DNA cassettes containing a recognition sequence for *Bsm*FI, a type IIS restriction endonuclease. Subsequent cleavage with the *Bsm*FI releases short (13–14 bp) but positionally defined sequences, referred to as tags, which are eventually concatenated into arrays and cloned into a plasmid vector for DNA sequencing. The power of the method, as its name implies, is that many tags can be read serially from each clone during the sequencing step which vastly increases throughput. Over 10 million cDNA tags have been analyzed by this method since it was first described, many of which are publicly available at (http://www.ncbi.nlm.nih.gov/SAGE).

In recent years, several new commercially available enzymes which cleave further into the DNA thereby increasing tag length, have superseded *Bsm*F1 since longer tags are particularly useful in characterizing expression patterns in the absence of complete genome sequence data, i.e., from "uncharted transcriptomes" and in designing primers to obtain full-cDNAs from transcripts whose tags are not currently present in RefSeq or similar expression databases. One very useful enzyme is *Mme*I which cleaves 20/18 bases past its non-palindromic (TCCRAC) recognition sequence *(20, 21)*. *Mme*I is now commercially available from NEB which has led to its use in numerous SAGE-type studies; typically referred to as Long SAGE due to the increase in tag length. In addition, detailed protocols are available on the web which explain many of the steps for preparation and analysis of SAGE libraries. A good starting point is the description of the I-SAGE™ Long kit (Catalog no. T5000-03) from Invitrogen (www.invitrogen.com)

3.4.1. Magnetic Beads Preparation and mRNA Capturing

1. Briefly vortex tube with Dynal oligo(dT) magnetic beads.
2. Remove 100 µl of bead suspension, place into clean 1.5 ml siliconized microcentrifuge tube and place tube on the magnetic holder. Beads will collect on the wall of the tube near the magnet.
3. Gently remove supernatant without disturbing the beads.
4. Wash the beads with 400 µl lysis-binding buffer from the Dynal Dynabeads mRNA Direct kit. To do this, remove tube from the magnet, wash beads off the wall using buffer, place the tube back on the magnet and wait for ~5 min to collect the beads.
5. Mix total platelet RNA (3–5 µg) with 200 µl lysis/binding buffer supplemented with 10 µg/ml glycogen.
6. Add RNA solution to the beads, incubate at 60°C for 5 min to help melt the secondary structure, cool slowly to room temperature with occasional mixing (once–twice per minute) to allow the poly (A) tracks in the mRNA to hybridize to the oligo dT.

7. Collect the beads by placing the tube back on the magnetic holder, carefully remove the supernatant, and store it at 4°C.
8. Wash beads twice using 400 μl wash buffer A, supplemented with 20 μg/ml glycogen.
9. Wash beads three times with 400 μl wash buffer B, supplemented with 20 μg/ml glycogen. Move beads to a fresh tube after the first wash.
10. Wash beads two times with 400 μl RT first strand buffer supplemented with 20 μg/ml glycogen and 2 μl SuperRNAseIn.

3.4.2. First Strand cDNA Synthesis

1. Re-suspend beads in 25 μl RT first strand buffer supplemented with 20 μg/ml glycogen and 2 μl SuperRNAseIn.
2. Incubate at 42°C for 2 min followed by incubation at 37°C for 2 min.
3. In a separate tube, prepare the following pre-mix for the reverse transcription reaction: 9 μl DEPC-treated water, 1 μl SuperRNAsin, 5 μl RT first strand buffer, 2.5 μl dNTP (10 mM each), 5 μl 0.1 M DTT, 2.5 μl SuperScript II reverse transcriptase.
4. Scrape beads off sides if necessary, add the pre-mix to the beads. Gently pipette up and down and incubate at 37°C for 1 h with gentle mixing.
5. Heat to 60°C for 3 min, incubate at 37°C for 2 min, and then add an additional 2 μl of reverse transcriptase.
6. Incubate at 37°C for an additional hour. Repeat one more time. Total volume of reverse transcriptase used for three cycles of cDNA synthesis equals 6.5 μl.
7. Collect beads and carefully remove RT first strand buffer.

3.4.3. Second Strand synthesis

1. In a separate tube, prepare a pre-mix: 254 μl DEPC-treated water, 70 μl 5× second strand buffer, 8 μl dNTP (10 mM each), 2.5 μl *Escherichia coli* DNA ligase, 10 μl *E.coli* DNA polymerase, 2.5 μl *E.coli* RNAse H, 3.5 μl glycogen.
2. Add pre-mix to the beads, incubate at 16°C overnight with gentle mixing for first 3 h.
3. Collect beads and wash six times with TEN/BSA buffer, using 500 μl per wash. For the first wash re-suspend in TEN/BSA, heat to 75°C for 10 min, and, slowly cool to room temperature.
4. Wash six more times, however, during the subsequent washes heating is not needed. After the fourth wash, transfer the beads with attached cDNA to a clean tube.

3.4.4. Digestion with NlaIII

1. Wash beads three times with 200 μl 1× New England Biolabs #4 buffer (optimal for *Nla*III).

2. Re-suspend beads in 200 µl 1× New England Biolabs #4 buffer, supplemented with 1× BSA and 4 mM spermidine $(HCl)_3$.
3. Add 2 µl *Nla*III (~10 U/µl).
4. Mix by pipetting up and down. Incubate at 37°C for 2 h.
5. Add an additional 2 µl *Nla*III and incubate for an additional 2 h at 37°C.
6. Capture beads, carefully remove as much supernatant as possible.
7. Wash beads with an additional 100 µl 1× NEB #4 buffer, pool and then precipitate the released DNA by adding 30 µl 2.5 M Na acetate, pH 5.2, and 1 µl glycogen. Mix well by vortexing.
8. Add 500 µl of cold 95% ethanol, vortex briefly, and incubate at –20°C for 2 h.
9. Centrifuge for 30 min at 12,000*g* in microcentrifuge.
10. Air-dry under vacuum for 3–5 min. Store beads in TEN/BSA buffer at 4°C.

3.4.5. Ligation of First Linker Cassette (Cassette A) Containing Restriction Site for MmeI Tagging Enzyme

The schema of *Mme*I restriction, cassette A and its ligation to cDNA is shown in **Fig. 16.4.**

1. Capture beads using magnetic holder, wash three times in 200 µl LoTE (10 mM Tris–HCl, pH 8.0; 0.1 mM EDTA), one time – in 200 µl 1× T4 DNA ligase buffer (Takara).
2. Make the following pre-mix for cassette A ligation: 38 µl LoTE, 5 µl 10× T4 DNA ligase buffer, 4 µl linker cassette A (mix of LS–LT primer and LS–LB primer, 10 pmoles/µl each)

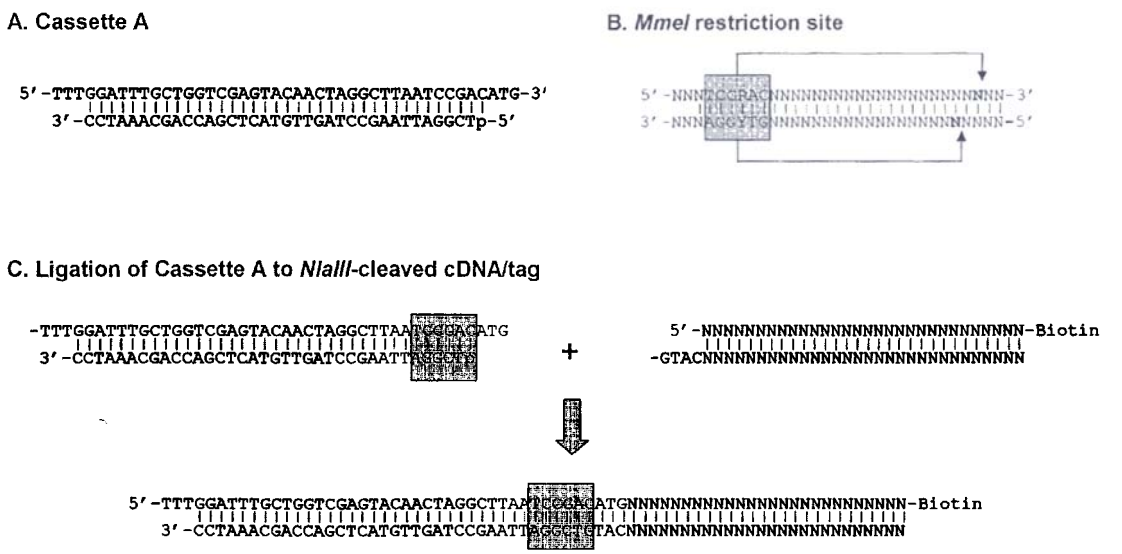


Fig. 16.4. Cassette A, *Mme*I restriction site and ligation of cassette A.

3. Heat pre-mix at 50°C for 2 min, cool to RT on the bench, and incubate at RT for 15 min.
4. Add 3 µl of T4 DNA ligase (~10 U). Mix well by pipetting.
5. Add to the beads and incubate at 16°C for 2 h with occasional gentle mixing.

3.4.6. MmeI Digestion to Generate Cassette A Ligated Beads

1. Wash beads six times with 400 µl TEN.
2. Capture beads, remove wash. Wash beads two times with 200 µl 1× *Mme*I buffer (NEB #4)
3. In a separate tube, make the following pre-mix: 85 µl ddH_2O (total volume will be 100 µl), 10 µl 10× *Mme*I buffer, 1 µl 100× SAM (4 mM – freshly prepare from 32 mM stock provided with the enzyme), 4 µl *Mme*I (~2 U/µl, 8 U total)
4. Incubate reaction at 37°C for 2 h and, mixing occasionally.
5. Collect beads using magnetic holder.
6. After *Mme*I digestion the cDNA/tags will be released into supernatant. Carefully remove as much supernatant as possible and transfer it to a fresh tube.
7. Rinse beads with 100 µl LoTE and combine with first supernatant.
8. To inactivate enzyme, extract supernatant with equal volume of TE-saturated phenol. Add phenol, vortex vigorously for 1 min and spin in microcentrifuge at top speed (~12,000*g*) for approximately 5 min at 4°C.
9. Carefully transfer upper aqueous phase to a fresh tube and extract with an equal volume of chloroform/IAA (24:1 v/v). Vortex briefly and spin in microcentrifuge at top speed for 15 min at 4°C.
10. Gently remove as much upper phase as possible, avoid disturbing the interphase. Place supernatant in a fresh tube.
11. Precipitate the released DNA/tags with ethanol. To 200 µl of supernatant, add 3 µl of GlycoBlue and 133 µl of 7.5 M ammonium acetate. Mix well and add 1 ml of cold 100% ethanol. Vortex and incubate at –70°C for 1 h or at –20°C overnight.
12. Spin at 12,000*g* for 30 min at 4°C. Wash twice with 75% ethanol, dry under vacuum for 5–10 min.

3.4.7. Ligation of Second Linker Cassette (Cassette B) Containing NlaIII Restriction Site

The schema of cassette B and its ligation to cassette A-tag is shown in **Fig. 16.5**. *Mme*I cuts 20–21/18–19 bp past its recognition site. Therefore cutting the linkered DNA in step 6, releases the linker and 17–18/15–16 bp immediately 3′ to the *Nla*III site. These sequences become the SAGE tags which are PCR amplified and ligated together to form ≥500 bp long concatemers prior to cloning and DNA sequencing. Because each clone contains multiple tags, sequencing throughput increases accordingly. Using the strategy described below, the two unique nucleotides at the 3′

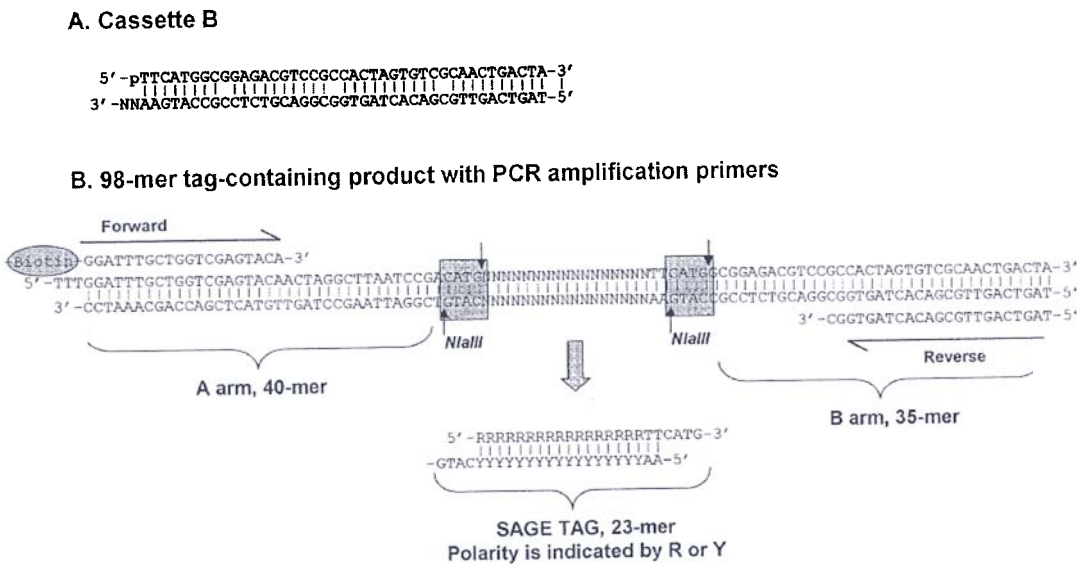


Fig. 16.5. Cassette B, PCR amplification product and concatamerized SAGE tags.

end of each tag are not lost. One strategy for capturing these nucleotides is based on the approach used in the process called TALEST (tandem arrayed ligation of expressed sequence tags) developed by Spinella et al. *(22)* and modified by our laboratory for our original GST protocol (6). It employs ligation with a 16-fold degenerate oligonucleotide (cassette B) to capture all the sequence information in the *Mme*I cleaved DNAs 3′ extensions.

1. Re-suspend DNA pellet (can be visible only as a blue dot because of GlycoBlue) in 29.5 µl TEsl.
2. Add 4 µl 10× T4 DNA ligase buffer and 3.5 µl linker cassette B (mix of LS–RT primer and LS–LB primer, 10 pmol/µl each).
3. Mix well, incubate at room temperature for 15 min, and then add 3 µl of T4 DNA ligase (~10 U).
4. Mix well by pipetting and incubate 2 h to overnight at 16°C.
5. Final product of ligation should be 98-mer with two *Nla*III restriction sites, as shown in **Fig. 16.5**, panel B.

3.4.8. PCR Amplification of cDNA Tags Using Biotinylated Primers

For this PCR, use biotinylated forward and reverse primers that corresponds to a portion of cassette A upper strand and to a portion of cassette B bottom strand (see **Fig. 16.4**).

1. Place clean microcentrifuge tube on ice and make PCR mix as follows: 371 µl ddH_2O, 50 µl 10× Promega buffer, 20 µl 50 mM $MgSO_4$, 15 µl dNTP (10 mM each), 20 µl 10

μM forward biotinyl-primer, 20 μl 10 μM reverse primer, 2.0 μl platinum *Taq* polymerase, 2.0 μl cDNA product from step 3.4.7. Store remaining cDNA product at –80°C for further analysis, if needed.
2. Mix well by brief vortexing, spin to remove traces of solution from walls of the tube.
3. Divide 500 μl PCR mix into ten small thin-wall PCR tubes, 50 μl per tube. Keep tubes on ice.
4. Run PCR using the following cycle parameters:
5. 95°C, 2 min 30 cycles with 95°C, 30 s;
58°C, 30 s; 72°C, 30 s 72°C,
4 min 10°C - hold

3.4.9. Optimizing PCR by Linear Amplification to Resolve Heteroduplexes (LARHD)

In our experience, standard PCR as above usually gives the expected ~94 bp product plus variable amounts of material migrating as a more slowly migrating diffuse band. As discussed in detail in Ref (GST paper), this material probably represents amplicons with perfectly base-paired cassette regions flanking SAGE tags with some miss paired bases. These products are then converted to completely duplex forms by a process, we termed LARHD.

1. Pool solution after PCR from ten tubes into one tube and mix well. Total volume should be 500 μl.
2. In a fresh tube, make the following pre-mix, total volume 1,250 μl: 682.5 μl ddH_2O, 125 μl 10× Promega buffer, 50 μl 50 mM MgSO_4, 37.5 μl dNTP (10 mM each), 50 μl 10 μM forward biotinyl-primer, 50 μl 10 μM reverse primer, 250 μl first round amplified tags from previous step (1/2 of the total volume), 5.0 μl platinum *Taq* Hi Fi polymerase.
3. Divide pre-mix into smaller thin-wall PCR tubes, keeping volume 50–100 μl per tube.
4. Run LARHD reaction (one cycle only): 95°C, 2 min 30 s; 58°C, 30 s; 72°C, 5 min; 10°C, hold.
5. Pool PCR products, saving 10 μl for gel analysis if desired. Products should be 94 bp in length, as shown in **Fig. 16.5.**

3.4.10. Exonuclease Digestion of Primers

1. To PCR products, add, 6.25 μl *E. coli* exonuclease I, 10 U/μl to 1,250 μl of LARHD products.
2. Incubate at 37°C for 30 min, remove and store 5 μl for gel analysis.
3. Perform phenol/chloroform purification as described in **Section 3.4.6.**
4. Wash lower (phenol) phase with a small volume of TEsl.
5. Pool and then precipitate in as many tubes as needed. One tube should contain: 270 μl sample, 30 μl 3 M Na acetate, pH 6.0, and 750 μl 100% ethanol (–20°C)
6. Place at –20°C for 20 min.

7. Spin down amplicons at 4°C, 30 min, 12,000*g*.
8. Redissolve in 0.3 M Na acetate, re-precipitate in one tube with 2.5 volumes of cold 100% ethanol. Place at –20°C for 20 min.
9. Centrifuge at 12,000*g* for 30 min at 4°C, remove supernatant, and carefully wash with 1 ml of cold 70% ethanol.
10. Spin down amplicons and air-dry.

3.4.11. NlaIII Digestion to Release SAGE Tags

1. Dissolve amplicons in 50 µl of 1× NEB buffer #4, supplemented with 1× BSA, and 4 mM spermidine HCl.
2. Add 4 µl of *Nla*III, mix well by pipetting, and incubate for 2 h at 37°C.
3. Add another 4 µl of *Nla*III and incubate for additional 2 h. (*see* **Fig.** 16.5 for schema of this restriction). This step releases the monomeric tags from the flanking cassettes used for PCR amplification.
4. Bring total volume to 200 µl with 1× NEB buffer #4 and mix well. Keep tube on ice.
5. Conduct phenol/chloroform purification as described in **Section 3.4.6** and precipitate with ethanol.

3.4.12. Gel-Purification of 23-mer SAGE Tags

In our hands, removing the biotinylated cassettes by capturing on streptavidin beads usually does not leave pure tags in the unbound fraction. Therefore we recommend loading the entire sample on a 10–12% non-denaturing polyacrylamide gel. It is important to electrophorese the sample slowly (100 V) until the bromophenol blue dye front reaches 1–2 cm from the bottom so that the tags are not heat denatured.

1. After electrophoresis, stain the gel in 0.5 µg/ml ethidium bromide plus 50 mM NaCl and visualize the bands under long wave length UV light. You should see a sharp 23–24 band below the cassette arms.
2. Using a clean razor blade, excise the tag band from each lane of the gel, place on a clean piece of parafilm, and then cut each gel piece into smaller pieces and place them into one or more 0.5-ml microcentrifuge tubes as needed that have had a hole pierced into the bottom by a 18-gauge needle.
3. Place each 0.5-ml microcentrifuge tube in a sterile 1.5-ml nonstick microcentrifuge tube and centrifuge at maximum speed for 2–3 min. The excised gel pieces will pass through the hole and be broken into much smaller pieces and be collected in the 1.5-ml microcentrifuge tube.
4. To each tube add 200–400 µl of LoTEsl plus 2.5 M ammonium acetate making sure that enough buffer is used to completely submerge the gel pieces.

5. Incubate at 37°C for several hours with occasional vortexing to elute the DNA.
6. Remove the eluates and pass them though a clean spin filter to remove any particles of polyacrylamide.
7. Combine the eluants and precipitate with ethanol. From this point, the procedure follows the standard SAGE methods for tag concatemerization, cloning into pZERO, and sequence analysis. The caveat, of course, is that our protocol generates runs of monomeric tags (**Fig. 16.5**), not the standard ditags that is the norm with other SAGE procedures.

4. Notes

1. All procedures of platelet purification should be performed at room temperature unless specified otherwise to avoid platelet activation. In our experience, platelet activation can be prevented during purification by using three inhibitors – EDTA or citrate, PGE, and aspirin. If platelets were activated during purification, the yield of total RNA is minimal and its quality is not good for microarray analysis, as estimated by hybridization to the test chip in Affymetrix protocol.
2. At this step, magnetic separation columns are used to capture any remaining $CD45^+$ cells (leukocytes) by positive selection. The amount of time needed for this step is somewhat unpredictable, but may take from 30 min to 2 h. Buffer flow may dramatically decrease or platelets may start to clog the columns. Although this procedure helps to remove traces of leukocyte contamination, it may be omitted with minimal effect on final platelet purity.
3. Avoid taking too much PRP for gel filtration since this may result in leukocyte contamination of the sample.
4. In small-scale platelet purification schema, platelets if necessary can be concentrated before gel filtration by centrifugation at 2,000*g* for 10 min and re-suspended 1–2 ml of the running buffer (HBMT) (**Section 3.3.2, step 1**). Add inhibitors to avoid platelet activation (see **Section 3.1.1**).
5. Trizol protocol for total RNA isolation allows simultaneous isolation of a protein fraction. This may be useful if further studies include immunodetection of protein of interest or validation of transcript profiling at the protein level. If necessary, harvest protein pellet before guanidine step exactly as described in Trizol reagent protocol and store it at –80°C for further analysis. Protein pellets can be stored for up to 6 months at –80°C without detectable degradation and be analyzed by the Western blotting.

Acknowledgments

This work was supported by NIH/NHLBI grants R21 HL076457 (D.V.G.) and HL086376 (W.F.B); Department of Defense grant MP048005 and a Targeted Research Award (D.V.G.) from Stony Brook University. Studies at the Brookhaven National Laboratory were supported by a Laboratory Directed Research and Development award (J.J.D.) and by the Offices of Biological and Environmental Research, and of Basic Energy Sciences (Division of Energy Biosciences) of the US Department of Energy.

References

1. Newman, P. J., Gorski, J., White, G. C., 2nd, et al. (1988) Enzymatic amplification of platelet-specific messenger RNA using the polymerase chain reaction. *J Clin Invest* 82, 739–743.
2. Kieffer, N., Guichard, J., Farcet, J. P., et al. (1987) Biosynthesis of major platelet proteins in human blood platelets. *Eur J Biochem* 164, 189–195.
3. Rinder, H., Schuster, J., Rinder, C., et al. (1998) Correlation of thrombosis with increased platelet turnover in thrombocytosis. *Blood* 91, 1288–1294.
4. Denis, M. M., Tolley, N. D., Bunting, M., et al. (2005) Escaping the nuclear confines: signal-dependent pre-mRNA splicing in anucleate platelets. *Cell* 122, 379–391.
5. Bugert, P., Dugrillon, A., Gunaydin, A., et al. (2003) Messenger RNA profiling of human platelets by microarray hybridization. *Thromb Haemost* 90, 738–748.
6. Gnatenko, D. V., Dunn, J. J., McCorkle, S. R., et al. (2003) Transcript profiling of human platelets using microarray and serial analysis of gene expression. *Blood* 101, 2285–2293.
7. McRedmond, J. P., Park, S. D., Reilly, D. F., et al. (2004) Integration of proteomics and genomics in platelets: a profile of platelet proteins and platelet-specific genes. *Mol Cell Proteomics* 3, 133–144.
8. Gnatenko, D. V., Cupit, L. D., Huang, E. C., et al. (2005) Platelets express steroidogenic 17beta-hydroxysteroid dehydrogenases. Distinct profiles predict the essential thrombocythemic phenotype. *Thromb Haemost* 94, 412–421.
9. Hillmann, A. G., Harmon, S., Park, S. D., et al. (2006) Comparative RNA expression analyses from small-scale, single-donor platelet samples. *J Thromb Haemost* 4, 349–356.
10. Shim, M. H., Hoover, A., Blake, N., et al. (2004) Gene expression profile of primary human CD34+CD38lo cells differentiating along the megakaryocyte lineage. *Exp Hematol* 32, 638–648.
11. Tenedini, E., Fagioli, M. E., Vianelli, N., et al. (2004) Gene expression profiling of normal and malignant CD34-derived megakaryocytic cells. *Blood* 104, 3126–3135.
12. Lockhart, D. J., Winzeler, E. A. (2000) Genomics, gene expression and DNA arrays. *Nature* 405, 827–836.
13. Hoheisel, J. D. (2006) Microarray technology: beyond transcript profiling and genotype analysis. *Nat Rev Genet* 7, 200–210.
14. Sausville, E. A., Holbeck, S. L. (2004) Transcription profiling of gene expression in drug discovery and development: the NCI experience. *Eur J Cancer* 40, 2544–2549.
15. Gnatenko, D. V., Perrotta, P. L., Bahou, W. F. (2006) Proteomic approaches to dissect platelet function: one-half of the story. *Blood* 108, 3983–3991.
16. Gnatenko, D. V., Bahou, W. F. (2006) Recent advances in platelet transcriptomics. *Transfus Med Hemother* 33, 217–226.
17. Butte, A. (2002) The use and analysis of microarray data. *Nat Rev Drug Discov* 1, 951–960.
18. Velculescu, V., Zhang, L., Vogelstein, B., et al. (1995) Serial analysis of gene expression. *Science* 270, 484–487.
19. Velculescu, V., Zhang, L., Zhou, W. et al. (1997) Characterization of the yeast transcriptome. *Cell* 88, 243–251.
20. Boyd, A. C., Charles, I. G., Keyte, J. W., et al. (1986) Isolation and computer-aided characterization of MmeI, a type II restriction endonuclease from Methylophilus methylotrophus. *Nucleic Acids Res* 14, 5255–5274.

21. Tucholski, J., Skowron, P. M., Podhajska, A. J. (1995) MmeI, a class-IIS restriction endonuclease: purification and characterization. *Gene* 157, 87–92.

22. Spinella, D. G., Bernardino, A. K., Redding, A. C., et al. (1999) Tandem arrayed ligation of expressed sequence tags (TALEST): a new method for generating global gene expression profiles. *Nucleic Acids Res* 27, e22.

Multiplexed genetic profiling of human blood platelets using fluorescent microspheres

Dmitri V. Gnatenko[1], Wei Zhu[2], Wadie F. Bahou[1,3]

[1]Department of Medicine, State University of New York, Stony Brook, New York, USA; [2]Department of Applied Mathematics and Statistics, State University of New York, Stony Brook, New York, USA; [3]Program in Genetics, State University of New York, Stony Brook, New York, USA

Summary

Human platelets have unique and reproducible mRNA profiles, with evidence for distinct profiles in haematopoietic stem cell disorders associated with thrombocytosis. Platelet transcript profiling is traditionally studied by microarray analysis, quantitative reverse transcription-PCR or serial analysis of gene expression, techniques that are labor- and technically-intensive. We have now applied a novel multiplex-based platform for quantitative transcript profiling of human platelets. Simultaneous quantification of 17 platelet transcripts was assayed using intact platelet-rich plasma or gel-filtered platelets lysed *in vitro*. Accurate and reproducible profiles could be obtained from as few as 5×10^7 platelets (a platelet mass corresponding to ~100 µl of whole blood), even for the low-abundant platelet transcripts. Correlation coefficients of this 17-member gene set to platelet Affymetrix microarrays were excellent ($r^2 = 0.949$, $p < 1 \times 10^{-10}$), with no correlation to in kind-derived leukocyte profiles, highlighting the cell-specificity of the platform. These data demonstrate that transcript multiplexing using fluorescent microspheres can be adapted for rapid molecular profiling using intact platelets (bypassing the need for RNA isolation methods), with potential applicability irrespective of baseline platelet counts.

Keywords

Platelets, transcript profiling, fluorescent microspheres, Luminex, mRNA

Thromb Haemost 2008; 100: 929–936

Introduction

Human blood platelets retain transcript expression profiles that are distinct from those of other blood cells (1–3). Recent studies highlight the uniqueness and complexity of the platelet transcriptome which is enriched with mitochondrial RNA (1, 4), while retaining a distinct evolutionary capacity for pre-mRNA splicing (5) and regulatory mRNA elements involved in mRNA localization and translational control (4). Furthermore, evidence from multiple laboratories has demonstrated that transcriptomic differences exist between normal and diseased platelets, such as those from patients with essential thrombocythemia (ET) (6), or patients with rare congenital platelet disorders associated with clinically-significant haemorrhage (7). More widespread applicability has also been demonstrated for patients with sickle-cell anemia or the larger cohort of patients with coronary artery disease (CAD) (8, 9). Sickle-cell disease was associated with global activation of platelet genes involved in arginine uptake and catabolism, presumably offering an explanation for limiting plasma arginine bioavailability for platelet nitric oxide synthesis. More recently, class prediction models have been proposed as a means of classifying platelet disorders (10, 11), with a long-term goal of extending these models for predicting platelet-associated myocardial infarction (MI)- or stroke-susceptibility risk.

Current methods of platelet transcript profiling (microarray, serial analysis of gene expression [SAGE], and quantitative real-time reverse-transcription PCR [Q-PCR]) represent well-established techniques that are poorly adapted to clinical laboratories due to the complexity of RNA isolation, experimental design, and subsequent data analysis (4, 12–14). Towards adapting transcript profiling technologies to clinical use, we applied a recently-developed platform for rapid and efficient platelet mRNA

Correspondence to:
Dmitri V. Gnatenko
Dept. of Medicine, HSCT15–040
State University of New York
Stony Brook, NY 11794–8151, USA
Tel.: +1 631 444 2059, Fax: +1 631 444 7530
E-mail: dgnatenk@notes.cc.sunysb.edu

Financial support:
This research was supported by grants HL49141 and HL086376 (National Institutes of Health), MP048005 (Department of Defense), NIH Center Grant MOI 10710–5 to the Stony Brook University General Clinical Research Center, and Targeted Research Award (Stony Brook University).

Received May 16, 2008
Accepted after minor revision August 11, 2008

Prepublished online October 13, 2008
doi:10.1160/TH08-05-0305

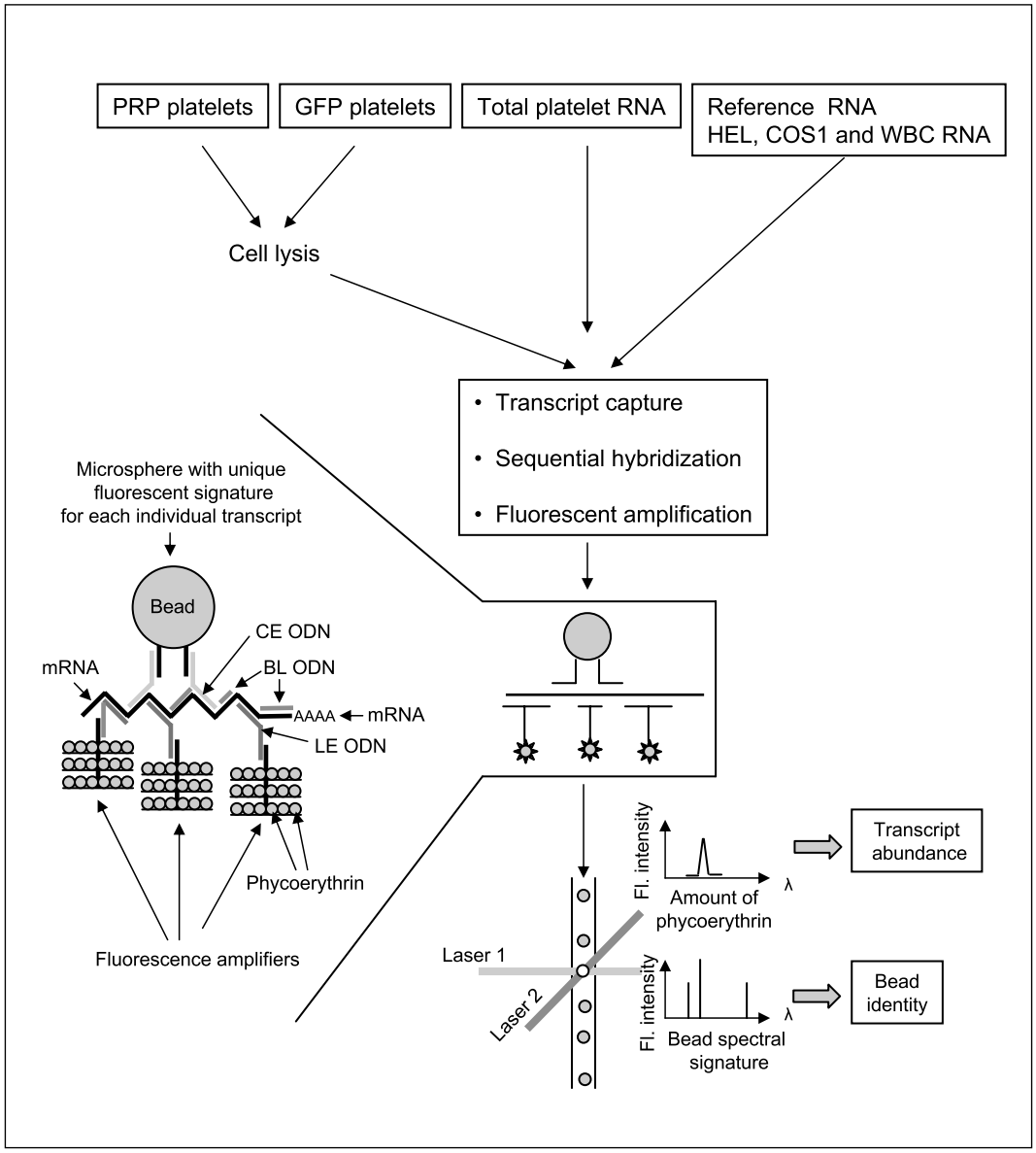


Figure 1: Schema of platelet multiplexing using microspheres. In this schema, transcripts are progressively captured, hybridized, and quantified using microspheres coupled to uniquely-designed oligonucleotides probes (ODN). The complete probe set for each individual transcript (26–30 oligonucleotides per transcript) includes three types of oligonucleotides: CE (capture extender) and LE (label extender) oligonucleotides are complementary to the mRNA of interest and contain unique "tails" that allow for coupling to microspheres (CE) or phycoerythrin (LE); BL (blocking) oligonucleotides fill in gaps, and are designed to stabilize captured mRNA molecules, and improve detection sensitivity and signal amplification. As individual target mRNA are captured to specific beads, signal amplification and detection occurs with the binding of streptavidin-conjugated fluorescent probes (phycoerythrin). As individual microspheres pass through the fluorescent detector, simultaneous read-outs link the gene ID (microsphere spectral signature) with transcript abundance since the fluorescent signal quantified from each microsphere is proportional to the number of captured mRNA transcripts. HEL – human erythroleukemia cells; WBC – leukocyte.

profiling. A multiplex, branched assay for quantitative transcript profiling was developed that provides efficient and reproducible transcript profiling from intact cells (15, 16). This approach bypasses the need for RNA isolation, allows for simultaneous determinations of multiple platelet transcripts from individual microtiter wells, and can be completed using intact platelets equivalent to those found in 0.1 ml of peripheral blood.

Methods

Platelet and RNA isolation

All studies were completed after approval from the Stony Brook University Committee on Research Involving Human Subjects (CORIHS). Blood (20 ml) was collected by venipuncture into sodium-citrated tubes (final concentration 0.4%), and used for isolation of pure leukocytes (by density-gradient centrifugation) or platelets, essentially as previously described (1). Briefly, platelet-rich plasma (PRP) was prepared by centrifugation for 3.5 minutes (min) at 1,800 rpm (~700 g) at 25°C; The upper 9/10 of PRP (~5 ml total) was subsequently harvested and supplemented with 0.1 μM prostaglandin E1 (PGE_1) and 10 mM EDTA, prior to RNA isolation or preparation of gel-filtered platelets (GFP). The latter were isolated over a Sepharose 2B column equilibrated with Hepes-buffered modified Tyrodes (HBMT: 10 mM HEPES [N-2-hydroxyethylpiperazine-N'-2ethanesulfonic acid], pH 7.45; 137 mM NaCl; 2.7 mM KCl; 0.4 mM NaH_2PO_4; 12 mM $NaHCO_3$; 0.2% BSA; 0.1% dextrose), in the presence of 0.1 μM prostaglandin E1 (PGE_1) and 10 mM EDTA. PRP and GFP were counted and pelleted by centrifugation (1,300 g; 7.5 min; 25°C) prior to analyses. Total platelet RNA was isolated from GFP three days prior to the experiment, and stored at –80°C. Total RNA was isolated using Trizol protocol as previously described (1), and resuspended in 50 μl of DEPC(diethylpyrocarbonate)-treated water containing 1 μl of RNAse Inhibitor (Ambion, Austin, TX, USA).

Table 1: List of platelet-expressed transcripts used for multiplexing.

Gene symbol	Description	RefSeq ID	Transcript abundance[1]
ACTB	Actin, beta	NM_001101	181279.9
HBA2	Human alpha-globin gene with flanks	NM_000517	85516.1
SH3BGRL3	SH3 domain binding glutamic acid-rich protein like 3	NM_031286	59992.5
TMEM111	30 kDa protein	NM_018447	53626.3
RBX1	Ring-box 1	NM_014248	28940.6
RPS20	Ribosomal Protein S20	NM_001023	12304.5
CTTN	Cortactin	NM_005231	9481.2
TPM1	Tropomyosin 1 (alpha)	NM_000366	8209.3
RPL32	Ribosomal Protein L32	NM_000994	6802.2
STOM	Erythrocyte membrane protein	NM_004099	6112.6
PKIG	Protein kinase inhibitor gamma	NM_007066	1759.8
PCSK6	Proprotein convertase subtilisin/ kexin type 6	NM_002570	1742.4
PDGFC	Platelet derived growth factor C	NM_016205	1419.4
WASF3	WAS protein family, member 3	NM_006646	653.0
TP53TG3	TP53TG3 protein	NM_015369	530.0
TMEM123	Pro-oncosis receptor inducing membrane injury gene	NM_052932	102.2
ACTR2	ARP2 actin-related protein 2 homolog	NM_005722	56.8

[1]Transript abundance is expressed as mean raw fluorescence intensity per transcript (rank-ordered from highest to lowest abundance), averaged from five platelet microarray analyses using the Affymetrix platform.

Probe and microsphere design

A multiplex gene expression analysis platform was developed for comparative transcript profiling using either intact cells or total cellular RNA. This branched DNA (bDNA) gene detection system is a sandwich nucleic hybridization assay that quantifies mRNA directly from cellular lysates by amplifying the reporter signal rather than target transcripts (15, 16). When coupled with the xMAP technology that assigns unique fluorescent signatures to individual microspheres, multiplexed quantification of up to 100 transcripts may be completed in parallel.

Seventeen platelet-expressed transcripts were chosen for multiplex profiling (Table 1). Transcripts were chosen to represent gene abundances at extreme ranges, as initially delineated using previously-isolated platelet expression profiles from five healthy control subjects (6). Microarray analysis using the Affymetrix platform had demonstrated that all 17 transcripts are expressed in normal human platelets, with a broad dynamic range of expression abundance that spanned nearly 4-logs. Three types of oligonucleotides were generated for each of the 17 mRNAs, collectively designed to optimize (i) mRNA capture, (ii) signal amplification, and (iii) mRNA stabilization. For each transcript, the probe set contained 26–30 individual oligonucleotides, uniquely designed to encompass ~500-base pair mRNA segment length. Each probe set typically included six capture extender oligonucleotides, 12 label extender oligonucleotides (enabling formation of 6 complete amplification molecules per mRNA molecule), and up to three blocking oligonucleotides (Fig. 1). Blocking oligonucleotides were designed to fill gaps in the probe set and to form complete double-stranded molecules, thereby stabilizing mRNA molecules while improving sensitivity of detection and signal amplification. Oligonucleotides were synthesized with 5'-NH_2-C6 linkers, and were covalently linked to carboxylated fluorescent-encoded microspheres (17), thereby providing unique microsphere-restricted signatures for each transcripts. For all microspheres and probe sets, coupling efficiencies were optimized and quality-controlled to minimize non-specific hybridizations (16). For some experiments, Affymetrix leukocyte expression profiles from three healthy volunteers were used for comparative analyses (6).

Platelet multiplexing

Platelet lysates (GFP or PRP) were prepared by solubilization into a cell lysis mixture (Panomics, Redwood City, CA, USA) supplemented with 50 µg/µl proteinase K, followed by a 30-min incubation at 65°C. After serial dilution (1:3 and 1:9) into the same lysis mixture, individual 80 µl aliquots were captured onto microspheres (2,000 microspheres of each type per assay) in a 100 µl reaction. For transcript profiling from intact platelets, the following number of platelets were used: [GFP – 5 x 10^7, 16 x 10^7, and 46 x 10^7 platelets; PRP – 6 x 10^7, 19 x 10^7, and 59 x 10^7 platelets.] For all experiments, hybridizations were completed in triplicate. Comparative analysis of total RNA was completed in the identical manner, using platelet, leukocyte, human erythroleukemia (HEL) cells or COS-1 total RNA as controls. Known

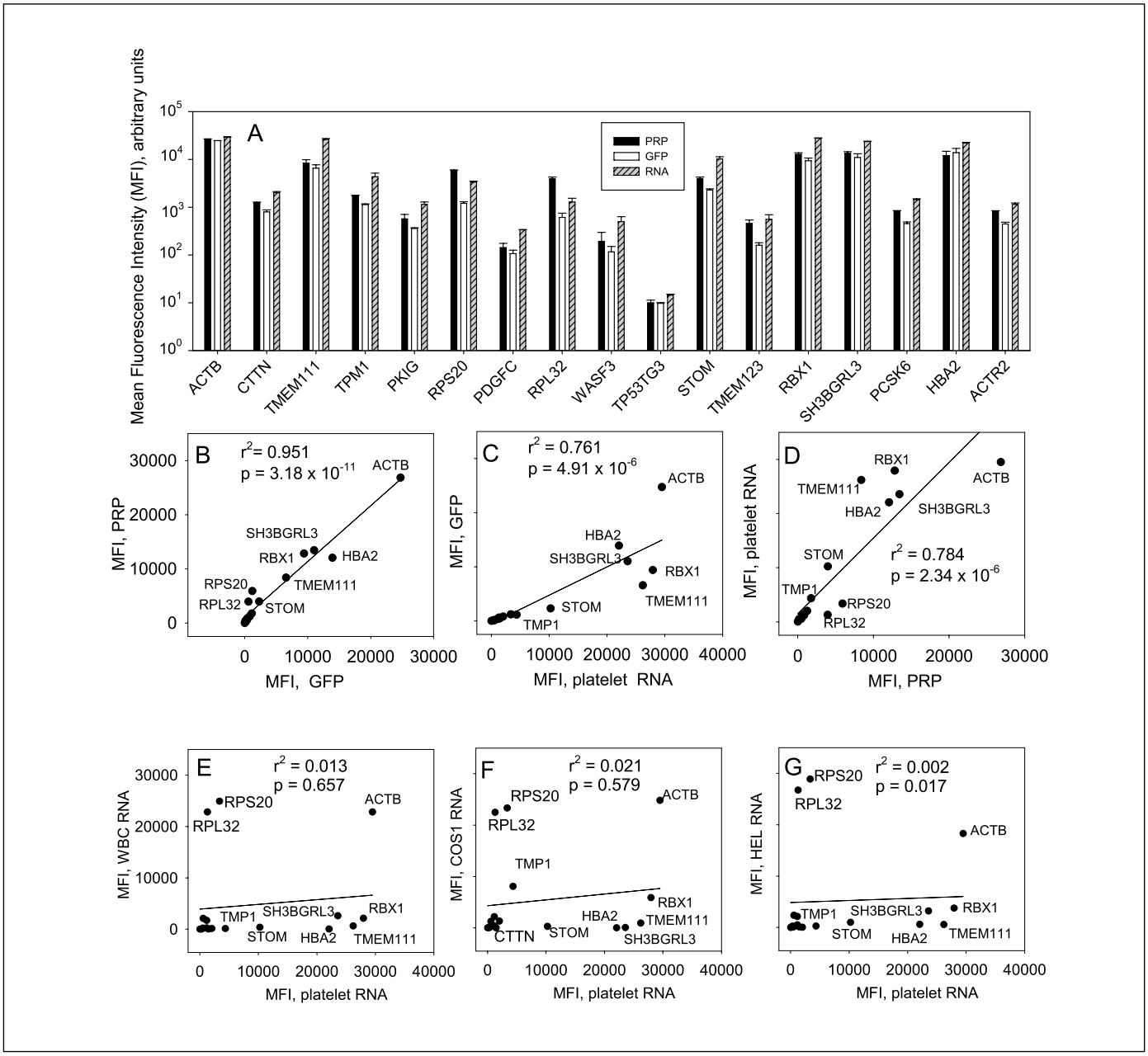


Figure 2: Comparative genetic multiplexing using distinct platelet mRNA sources. A) The relative abundance of 17 transcripts was quantified using purified platelet RNA (840 ng/well) and intact platelet fractions estimated to contain comparable amounts of RNA (6). Intact platelet samples included GFP platelets (1.6 x 10^8/well) and platelet-rich plasma (PRP, 1.9 x 10^8/well). Results are expressed as means ± standard deviations from triplicate wells, and are representative of one complete set of experiments. B – G) Regression analyses comparing transcript profiles obtained from PRP, GFP, or total platelet RNA demonstrate excellent correlation (B–D), with poor correlation coefficients when compared to non-platelet RNA sources (leukocytes, HEL, COS-1 cells (all 333 ng/well) (E – G).

dilutions of universal human reference RNA (Stratagene, Cedar Creek, TX, USA) were used as additional controls as necessary.

After sealing individual wells, hybridizations were allowed to proceed for 16–18 hours overnight at 54°C. Following the overnight capture of the target mRNAs, microspheres were transferred onto 0.45 µm filters (Millipore, Billerica, MA, USA), washed, and sequentially hybridized at 50°C with the bDNA amplifier and 5'-dT(biotin)-conjugated label probes. Unbound material was washed from microspheres using a vacuum manifold and 0.1 X SSC/0.03% lithium lauryl sulfate, followed by a 30-min incubation at 25°C using streptavidin-conjugated R-phycoerythrin (SAPE). The microspheres were then washed to remove unbound SAPE, followed by read-outs using a BioPlex reader (Bio-Rad, Hercules, CA, USA) calibrated to the high sensitivity mode.

Data analyses

Relative transcript abundance for each individual gene was determined using the raw fluorescent signal intensities that were

obtained from 100 microspheres, and are reported as the mean fluorescent intensity (MFI). The identity (ID) of each target gene was defined and linked to a specific microsphere by design; during the experimental phase, individual microsphere identifiers were read by the instrument in tandem with quantitative fluorescence signals, thereby providing simultaneous read-outs of gene ID and transcript abundance. Raw fluorescent data were exported to Microsoft Excel software for data analysis. For all determinations, background signals were established in the absence of target RNAs, and were subtracted from signals derived in the presence of RNAs. The sensitivity of the assay for individual target RNAs was delineated by determining the limit of detection, empirically defined as the target concentration at which the signal is three standard deviations above background. For all experiments, statistical significance was determined by analysis of variance (1- or 2-way ANOVA), while correlation coefficients were established using regression analysis. For all biological comparisons, $p < 0.05$ was used to establish statistical significance.

Results and discussion

Comparative multiplexing using distinct platelet fractions

The list of 17 transcripts known to be expressed in human platelets using both Affymetrix and in-house microarray platforms is delineated in Table 1. As established using the Affymetrix HG-U95Av2 (~12,600-probe set) gene arrays (6), the transcripts encompassing the 17-gene dataset represented a broad dynamic abundance range (4-logs), thereby providing an optimal dataset for comparative multiplexed analyses. Initially, relative abundance was studied using two distinct platelet purification fractions (PRP or gel-purified platelets [GFP]). GFP or PRP transcript quantifications were completed using cells lysed *in vitro*, and were compared to transcript studies using purified platelet RNA analyzed in parallel. Despite the broad range of relative expression, all transcripts were detected using the microspheres (Fig. 2A). For the majority of the transcripts, the signal from total platelet RNA was slightly higher than from GFP or PRP, although the correlations coefficients comparing each of the starting materials (pure platelet RNA, GFP, PRP) were excellent (Fig. 2B-D). For five of the transcripts (*ACTB*, *TMEM*, *RBX1*, *SH3BGRL3* and *HBA2*), the MFI using total platelet RNA approached the plateau, indicating that the amount of total RNA taken for the initial analysis may be excessive (*see below*). Overall, the standard error bars were quite small, demonstrating excellent reproducibility for both high- and low-abundant transcripts, using any of the RNA sources. In contrast, total RNAs from three other cell types (leukocytes, COS1, and HEL cells) demonstrated no significant correlation to platelet profiles, establishing the specificity of the multiplexing reactions (Fig. 2E-G).

Sensitivity studies using small platelet volumes

To address the sensitivity of transcript profiling using the microspheres, studies were repeated using varying amounts of GFP or PRP (Fig. 3A). Signals for 16/17 transcripts were reliably detected from as few as 5×10^7 platelets (a platelet mass typically found in ~100 µl of blood); only the *TP53TG3* transcript was at the lower limit of detection using this platelet number. For the highly abundant β-actin gene (*ACTB*), the fluorescent signal reached a plateau at 4.8×10^8 platelets, confirming microsphere saturation using platelet numbers found in ~ 1 ml of blood. This high level of *ACTB* is in excellent agreement with previously reported (6) microarray data (see Table 1). The signal for one other transcript (*HBA2*) was also non-linear over the dose-range, confirming its early fluorescence saturation at lower platelet starting numbers. Highly abundant hemoglobin transcripts have been detected in most platelet microarray studies conducted to date (1–3, 18), results reaffirmed by our multiplex assay. The remaining 15 transcript curves demonstrated excellent linearity, suggesting accurate detection over varying platelet numbers. Indeed, analysis of gene expression at different sample dilutions demonstrated excellent correlation using either PRP or GFP (Fig. 3B and C), even for low-abundant transcripts. These data also suggest that the optimal number of platelets for microsphere-based quantification of the 17-gene dataset is between $1.5 – 2 \times 10^8$ platelets per well, based on signal intensity and saturation plateau. This, however, may differ as other gene sets are developed and optimized for analysis (16). While these data document the high sensitivity and reproducibility of transcript detection over a 1-log range of starting platelets, they also establish that patients with profound thrombocytopenia ($< 20 \times 10^6$ platelets/ml) may also be studied using blood volumes typically obtained from routine phlebotomy (<10 ml).

Comparative abundance determinations using Affymetrix microarrays

To validate expression levels, multiplexed profiles obtained from 1.6×10^8 GFP platelets were compared to a normal platelet transcriptome database comprised of five highly purified apheresis platelet microarrays; these data were obtained using non-amplified platelet RNA, providing for technically comparable studies using multiplexing and Affymetrix platforms(6). Regression analysis (Fig. 4A) demonstrated excellent concordance between two platforms ($r^2 = 0.949$, $p < 1 \times 10^{-10}$). For comparison, expression levels of these transcripts were compared to normal leukocyte Affymetrix microarrays ($N = 3$) (6), demonstrating no significant correlation and highlighting the specificity of these platelet expression profiles (Fig. 4B). In contrast, the multiplexed leukocyte profiles demonstrated excellent concordance to their Affymetrix expression levels ($r^2 = 0.954$, $p < 1 \times 10^{-10}$), re-affirming the cell-specificity of the platform.

In summary, we have demonstrated that a multiplex assay can be applied for rapid quantification of multiple platelet transcripts. The assay can be completed using intact platelets lysed *in vitro*, and is therefore readily adaptable for high-throughput screening, bypassing the need for sophisticated RNA isolation techniques. Although transcript abundance determination is possible using quantitative (real-time) reverse-transcription PCR (qRT-PCR), the latter approach has minimal multiplexing capacity and is especially sensitive to DNA contamination and inter-sample differences in amplification efficiency (19). Furthermore, since this approach analyzes signal

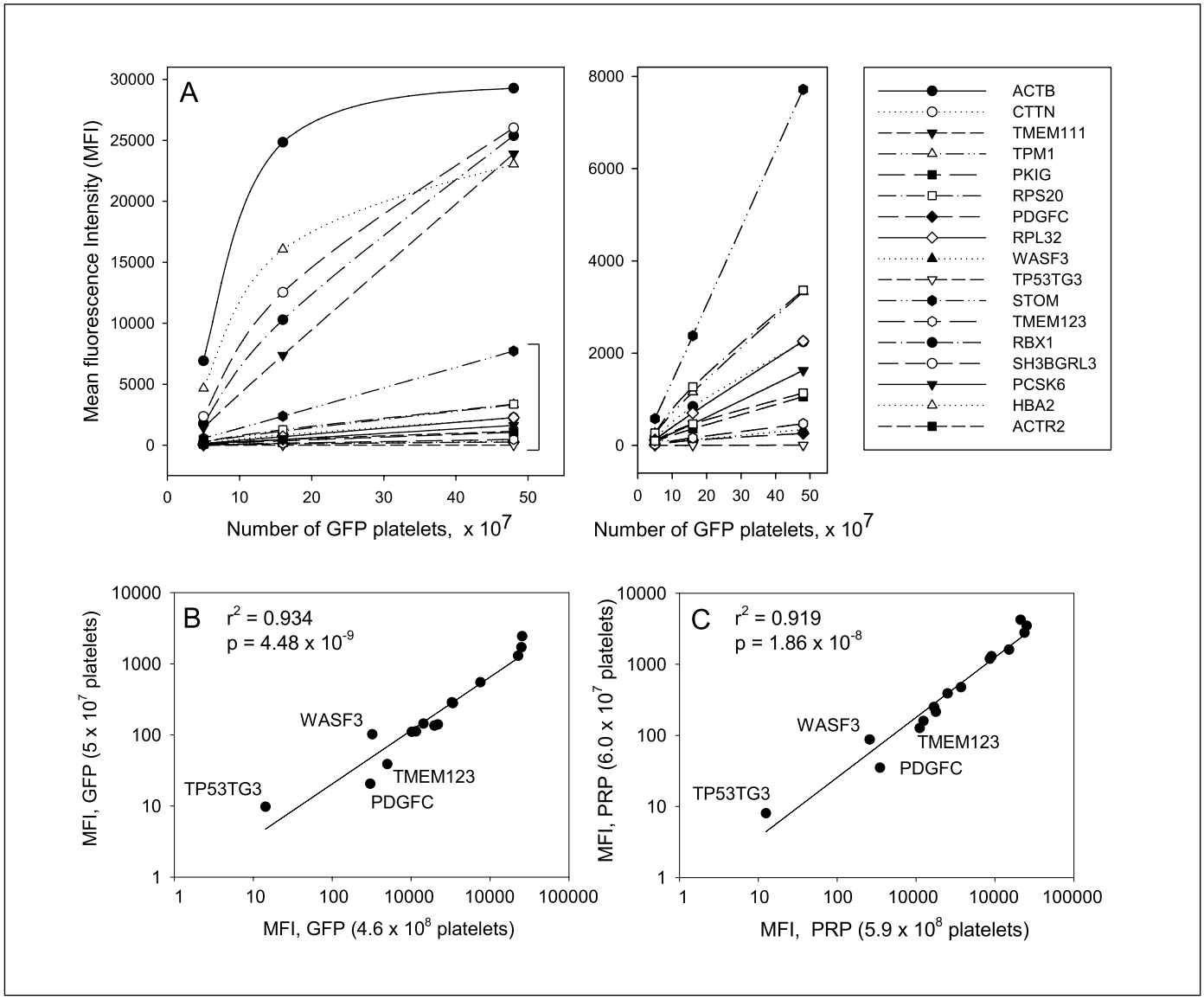


Figure 3: Linearity, sensitivity and reproducibility of platelet multiplexed analyses. A) Varying amounts of GFP (5 X 10^7, 16 X 10^7, 48 X 10^7) were directly aliquoted and lysed *in vitro*, followed by platelet multiplexing. The right plot shows an expanded scale for lower abundant transcripts. Note the clear linearity of the assay with the exception of highly abundant *ACTB* and *HBA2* genes. B, C) Correlation coefficients were determined for either GFP (B), or PRP (C), comparing multiplexed studies using high or low platelet numbers (delineated on axis). For these plots, *ACTB* and *HBA2* transcripts were removed from the analysis because of their saturation using high platelet number of platelets. Note the high correlation of expression for both high- and low-abundant transcripts using either platelet fraction.

amplification (as opposed to target amplification methodologies using qRT-PCR), there is the theoretical advantage for minimizing other technical variables during complex processing steps. Finally, multiplexing using fluorescent microspheres excludes mRNA and oligonucleotide handling, while minimizing the cumbersome processes of data analysis and interpretation.

Our data confirm that comparable results may be obtained using either GFP or PRP. Dose-response curves using different starting platelet numbers demonstrate consistent transcript expression measurements, and reproducible profiling can be completed using as few as 5 X 10^7 platelets. Theoretically, the assay can be completed using a platelet mass readily obtained from ~100 µl of whole blood, with persistent applicability to thrombocytopenic patients using physiologically relevant phlebotomy (10 – 20 ml). The expression patterns for the selected 17-member gene set strongly correlate with those obtained using a completely different platform and quantification algorithms (Affymetrix), further confirming its validity.

In the classification of leukemia by gene profiling, 50 genes were initially used, although a more select list (10 genes) displayed comparable discriminatory power (20). Similarly, a discrete list of 21 genes effectively predicted recurrence in subsets of breast cancer patients (21). While these and other

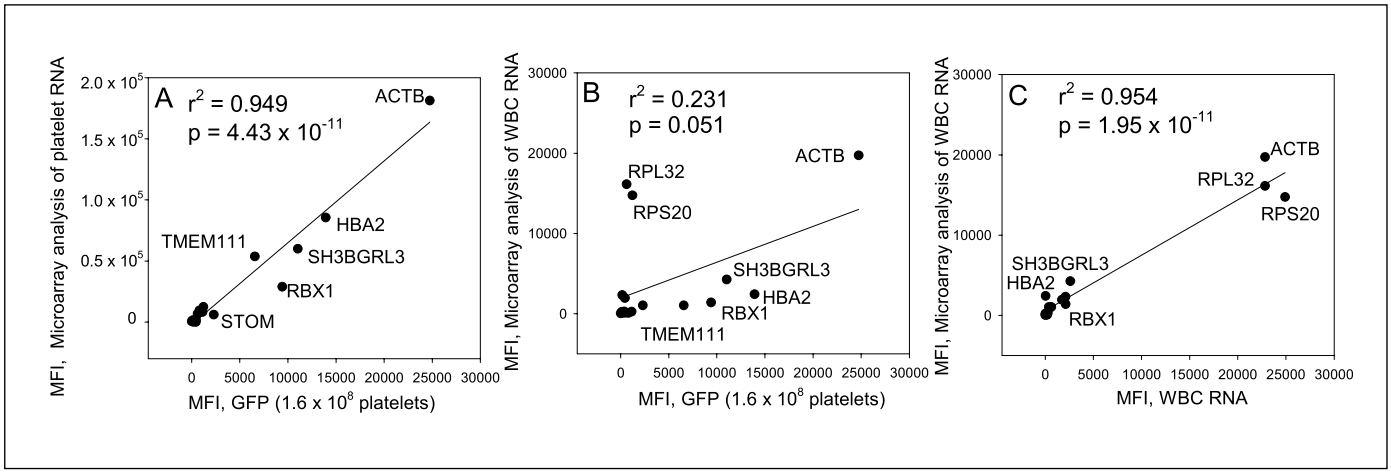


Figure 4: Platelet multiplex correlation to Affymetrix HG-U95Av2 arrays. A) Multiplexed expression values were obtained using GFP platelets (1.6 x 10^8) obtained from triplicate wells, while the corresponding 17-transcript dataset was extracted and averaged from platelet microarray analysis of five healthy donors, all obtained using the HG-U95Av2 Affymetrix microarray chip (6). Note the excellent correlation using two distinct platforms for the study of platelet gene expression. B) Multiplexed platelet transcript profile demonstrates no correlation with leukocyte expression profile (leukocyte microarray profile was obtained from the analysis of three healthy donors using the HG-U95Av2 microarray chip). C) Multiplexed leukocyte transcript profile demonstrates excellent concordance with leukocyte microarray data.

studies have focused on development of class prediction models for cancer diagnosis and/or prognosis, preliminary data from this laboratory have established that comparable approaches can be used to assign (predict) platelet phenotypic class using microarray expression profiling (10). By determining platelet microarray profiles in nearly 100 normal and thrombocytotic patients, we have identified an 11-member gene list that can effectively discriminate among normal, reactive and essential thrombocytotic platelets. Although these results will require ongoing confirmation, they strongly suggest that a small biomarker subset can be used for platelet diagnostics, an issue of considerable relevance for distinguishing among various causes of thrombocytosis as an example (22). Since the current platform has the capacity to simultaneously quantify expression data from over 40 genes (16), these data highlight the potential applicability of this strategy as an adaptable platform for molecular platelet classification.

What is known about this topic?

– Platelet transcript profile is unique, it can potentially be used to study molecular mechanisms of platelet diseases and for diagnostics.
– Traditional approaches to transcript profiling (microarray, SAGE and quantitative PCR) require relatively large amount of high quality RNA, which is difficult to achieve in case of platelets.

What does this paper add?

– We applied novel fluorescent microspheres-based technology to the profiling of selected transcripts from as few as 5 x 10^7 human platelets.
– Main advantages of this approach are that (i) it excludes any RNA manipulation, (ii) allows multiplexing and (iii) is performed in 96-well plate automatically.
– Fluorescent microspheres-based technology represents a novel perspective approach to platelet transcript profiling due to high sensitivity and accuracy. It can be accommodated for use in clinical laboratories.

References

1. Gnatenko DV, Dunn JJ, McCorkle SR, et al. Transcript profiling of human platelets using microarray and serial analysis of gene expression. Blood 2003; 101: 2285–2293.

2. Bugert P, Dugrillon A, Gunaydin A, et al. Messenger RNA profiling of human platelets by microarray hybridization. Thromb Haemost 2003; 90: 738–748.

3. McRedmond JP, Park SD, Reilly DF, et al. Integration of proteomics and genomics in platelets: a profile of platelet proteins and platelet-specific genes. Mol Cell Proteomics 2004; 3: 133–144.

4. Dittrich M, Birschmann I, Pfrang J, et al. Analysis of SAGE data in human platelets: features of the transcriptome in an anucleate cell. Thromb Haemost 2006; 95: 643–651.

5. Denis MM, Tolley ND, Bunting M, et al. Escaping the nuclear confines: signal-dependent pre-mRNA splicing in anucleate platelets. Cell 2005; 122: 379–391.

6. Gnatenko DV, Cupit LD, Huang EC, et al. Platelets express steroidogenic 17beta-hydroxysteroid dehydrogenases. Distinct profiles predict the essential thrombocythemic phenotype. Thromb Haemost 2005; 94: 412–421.

7. Sun L, Gorospe JR, Hoffman EP, et al. Decreased platelet expression of myosin regulatory light chain polypeptide (MYL9) and other genes with platelet dysfunction and CBFA2/RUNX1 mutation: insights from platelet expression profiling. J Thromb Haemost 2007; 5: 146–154.

8. Raghavachari N, Xu X, Harris A, et al. Amplified expression profiling of platelet transcriptome reveals changes in arginine metabolic pathways in patients

with sickle cell disease. Circulation 2007; 115: 1551–1562.
9. Healy AM, Pickard MD, Pradhan AD, et al. Platelet expression profiling and clinical validation of myeloid-related protein-14 as a novel determinant of cardiovascular events. Circulation 2006; 113: 2278–2284.
10. Gnatenko D, Zhu W, Dhundale A, et al. Class prediction model of thrombocytosis using gene profiling. Blood 2007; 110: 38A.
11. Gnatenko D, Perrotta P, Ji C, et al. Platelet gene/protein expression analyses using an integrated platform. Blood 2007; 110: 1065A.
12. Gnatenko DV, Bahou WF. Recent advances in platelet transcriptomics. Transfus Med Hemother 2006; 33: 217–226.
13. Gnatenko DV, Perrotta PL, Bahou WF. Proteomic approaches to dissect platelet function: Half the story. Blood 2006; 108: 3983–3991.
14. Senzel L, Gnatenko D, Bahou W. Platelet transcriptome and cardiovascular disease. Future Cardiol 2007; 3: 391–398.
15. Zheng Z, Luo Y, McMaster GK. Sensitive and quantitative measurement of gene expression directly from a small amount of whole blood. Clin Chem 2006; 52: 1294–1302.
16. Flagella M, Bui S, Zheng Z, et al. A multiplex branched DNA assay for parallel quantitative gene expression profiling. Anal Biochem 2006; 352: 50–60.
17. Yang L, Tran DK, Wang X. BADGE, Beads Array for the Detection of Gene Expression, a high-throughput diagnostic bioassay. Genome Res 2001; 11: 1888–1898.
18. Macaulay IC, Tijssen MR, Thijssen-Timmer DC, et al. Comparative gene expression profiling of in vitro differentiated megakaryocytes and erythroblasts identifies novel activatory and inhibitory platelet membrane proteins. Blood 2007; 109: 3260–3269.
19. Bustin SA, Nolan T. Pitfalls of quantitative real-time reverse-transcription polymerase chain reaction. J Biomol Tech 2004; 15: 155–166.
20. Golub TR, Slonim DK, Tamayo P, et al. Molecular classification of cancer: class discovery and class prediction by gene expression monitoring. Science 1999; 286: 531–537.
21. Paik S, Shak S, Tang G, et al. A multigene assay to predict recurrence of tamoxifen-treated, node-negative breast cancer. N Engl J Med 2004; 351: 2817–2826.
22. Schafer AI. Thrombocytosis. N Engl J Med 2004; 350: 1211–1219.

The NEW ENGLAND JOURNAL of MEDICINE

Please review the Supplemental Files folder to review documents not compiled in the PDF.

Molecular Classification and Prediction Models of Thrombocytosis

Journal:	*New England Journal of Medicine*
Manuscript ID:	draft
Article Type:	Original Article
Date Submitted by the Author:	n/a
Complete List of Authors:	Gnatenko, Dmitri; State University of New York at Stony Brook, Department of Medicine Zhu, Wei; State University of New York at Stony Brook, Department of Applied Mathematics and Statistics Xu, Xiao; State University of New York at Stony Brook, Department of Applied Mathematics and Statistics Monaghan, Melissa; State University of new York at Stony Brook, Center for Biotechnology Zarrabi, Mohammad; State University of New York at Stony Brook, Department of Medicine Kim, Christie; State University of New York at Stony Brook Dhundale, Anil; Long Island High Technology Incubator Bahou, Wadie; State University of New York at Stony Brook, School of Medicine
Key Words:	
Abstract:	BACKGROUND. Hematologic criteria for distinguishing among the various causes of thrombocytosis are limited in their capacity to delineate clonal (essential thrombocythemia; ET) from non-clonal (reactive thrombocytosis; RT) etiologies. We have developed algorithms that can predict phenotypic class using distinct genetic biomarker subsets. METHODS. Using a focused microarray gene chip, we studied the platelet genetic profiles of 95 patients (48 healthy controls, 23 RT, 24 ET [11 contained the Jak2V617F mutant allele; 13 were homozygous normal]). Gender differences in expression profiles were analyzed by cohort. Linear discriminant analysis with cross-validation was used to identify subsets that segregated phenotypes based on microarray profiles. Cross-platform consistency of platelet gene profiles was validated using quantitative real-time reverse-transcription polymerase chain reaction (qRT-PCR). Class prediction algorithms using the genetic biomarkers were developed to assign phenotypic class between the thrombocytosis cohorts, and by Jak2 genotype (within the ET cohort). RESULTS. Gender differences were rare in normal and ET cohorts (<1% of genes), but were male-skewed for ~3% of RT genes.

	An 11-biomarker gene subset was identified that discriminated among the three cohorts with 86.3% accuracy. Two-way class prediction (ET vs. RT) was successful in 93.6% of patients using microarray profiling and 90.0% of patients using qRT-PCR. A 4-biomarker gene subset predicted Jak2-wild type ET with >90% accuracy using either microarray or qRT-PCR profiling, with less robust class prediction in Jak2V617F-containing ET patients. CONCLUSIONS. Distinct genetic biomarker subsets can be developed to predict thrombocytosis class and Jak2V617F-negative ET using routine phlebotomy.

Molecular classification and prediction models of thrombocytosis

Dmitri V. Gnatenko[1]
Wei Zhu[2]
Xiao Xu[2]
Melissa Monaghan[3]
Mohammad H. Zarrabi[1,4]
Christie Kim[1]
Anil Dhundale[5]
Wadie F. Bahou[1,6]

[1]Department of Medicine, State University of New York, Stony Brook, NY 11794

[2]Department of Applied Mathematics and Statistics, State University of New York, Stony Brook

[3]Center for Biotechnology, State University of New York, Stony Brook

[4]Veteran's Administration Medical Center, Northport, NY 11768

[5]Long Island High Technology Incubator, Stony Brook, NY 11790

[6]Program in Genetics, State University of New York at Stony Brook

AUTHOR CONTACT: EMAIL: wbahou@notes.cc.sunysb.edu
Phone: 631-444-2059
FAX: 631-444-7530

MAILING ADDRESS: Department of Medicine, HSCT15-040, State University of New York at Stony Brook, 11794-8151

ACKNOWLEDGEMENTS

This work was supported by grants HL49141, HL53665, HL76457, Department of Defense #MPO48005, a Targeted Research Award from Stony Brook University, and NIH Center grant MO1 10710-5 to the University Hospital General Clinical Research Center.

ABSTRACT

BACKGROUND. Hematologic criteria for distinguishing among the various causes of thrombocytosis are limited in their capacity to delineate clonal (essential thrombocythemia; ET) from non-clonal (reactive thrombocytosis; RT) etiologies. We have developed algorithms that can predict phenotypic class using distinct genetic biomarker subsets.

METHODS. Using a focused microarray gene chip, we studied the platelet genetic profiles of 95 patients (48 healthy controls, 23 RT, 24 ET [11 contained the Jak2V^{617}F mutant allele; 13 were homozygous normal]). Gender differences in expression profiles were analyzed by cohort. Linear discriminant analysis with cross-validation was used to identify subsets that segregated phenotypes based on microarray profiles. Cross-platform consistency of platelet gene profiles was validated using quantitative real-time reverse-transcription polymerase chain reaction (qRT-PCR). Class prediction algorithms using the genetic biomarkers were developed to assign phenotypic class between the thrombocytosis cohorts, and by Jak2 genotype (within the ET cohort).

RESULTS. Gender differences were rare in normal and ET cohorts (<1% of genes), but were male-skewed for ~3% of RT genes. An 11-biomarker gene subset was identified that discriminated among the three cohorts with 86.3% accuracy. Two-way class prediction (ET *vs.* RT) was successful in 93.6% of patients using microarray profiling and 90.0% of patients using qRT-PCR. A 4-biomarker gene subset predicted Jak2-wild type ET with >90% accuracy using either microarray or qRT-PCR profiling, with less robust class prediction in Jak2V^{617}F-containing ET patients.

CONCLUSIONS. Distinct genetic biomarker subsets can be developed to predict thrombocytosis class and Jak2V^{617}F-negative ET using routine phlebotomy.

INTRODUCTION

Platelets mediate the initial first-step in hemostasis while simultaneously providing the negatively charged phospholipid surface required for contact phase-mediated propagation of the coagulation cascade. Despite these key functions, molecular defects causally implicated in platelet-associated bleeding or thrombotic risk are largely unknown, best-characterized by loss of glycoproteins (GP) IIb/IIIa [$\alpha_{IIb}\beta_3$; Glanzmann thrombasthenia] or the GPIb-IX-V complex [Bernard-Soulier syndrome].[1] Such limitations highlight the relevance of identifying gene/protein biomarkers that may be linked to clinically-important platelet phenotypes. Although platelets retain megakaryocyte-derived mRNA, the platelet transcriptome is less complex than that of nucleated cells,[2, 3] with evolution of unique adaptive molecular signals for maintenance of genetic and protein diversity.[4] Quiescent platelets generally display minimal translational activity, although maximally-activated platelets retain the capacity for protein synthesis.[5] Newly-formed "reticulated" platelets retain larger quantities of mRNAs and have been associated with enhanced thrombotic risk in patients with thrombocytosis,[6] although it remains unknown if this risk is due to specific changes in gene/protein subsets.

We have now applied gene expression profiling to identify biomarkers that can be used for class prediction models of thrombocytosis.[7] The differentiation of clonal from reactive thrombocytosis (RT) has important diagnostic and therapeutic implications since thrombohemorrhagic complications arising in RT are unusual,[8] in contrast to frequent events in patients with clonal disorders such as essential thrombocythemia (ET).[9-11] ET is a myeloproliferative disorder (MPD) subtype microscopically indistinguishable (unless accompanied by macrothrombocytosis) from the larger subset of non-clonal, thrombocytotic disorders associated with a wide array of human diseases. Recent data have identified an

activating *Janus kinase 2* mutation (Jak2V^{617}F) in patients with polycythemia rubra vera (PV)[12-14], although it appears with lower frequency in ET patients. Although an etiology for thrombocytosis is evident in many patients, its association with occult malignancies -- coupled with the fact that ET remains a diagnosis of exclusion[15] -- support the need for well-defined diagnostic criteria. Discriminatory class prediction models using gene expression profiling have been developed for human malignancies.[16-18] We now demonstrate that a comparable strategy can be developed for assigning class in patients with thrombocytosis.

METHODS

A. Patient recruitment. Patients were randomly enrolled from the larger pool of patients referred to the Division of Hematology for evaluation of thrombocytosis. All subjects provided informed consent for an IRB (Institutional Review Board)-approved protocol completed in conjunction with the Stony Brook University General Clinical Research Center. Standard hematological criteria were applied for the diagnosis of essential thrombocythemia or reactive thrombocytosis.[19, 20] Both sex- and age-distribution paralleled prevalence figures for ET with a M:F ratio of 1:2; age at diagnosis ranged from 23-78 years. Platelet counts at the time of blood isolation ranged from normal (reflecting treatment) to 1,724,000/μL; utilization of platelet-lowering drugs (i.e. hydroxyurea, analgrelide, or untreated) was recorded for individual patients at the time of platelet isolation and purification (**Table 1**).

B. Molecular studies. Leukocytes and gel-filtered platelets were isolated from peripheral blood (20 mL) as previously described;[2] the final platelet-enriched product contained no more than 3-5 leukocytes per 1 X 10^5 platelets. High-quality platelet RNA was isolated using Trizol,[21] and platelet mRNA quantification and integrity were established using an Agilent 2100 Bioanalyzer; mean platelet RNA concentrations among the three groups were comparable, ranging from ~0.3 - 1.0 fg/platelet. High molecular weight DNA was used as the source for genomic Jak2V^{617}F (exon 12, 1849$^{G\rightarrow T}$ transversion) genotyping,[12-14] while platelet mRNA was used for cellular genotyping. Mutational screening was completed using both pyrosequence and dideoxy sequence analyses of PCR-amplified fragments. Samples were defined as Jak2V^{617}F-positive if the mutant allele was detected in >5% of the nucleic acid pool.

Confirmatory studies of platelet gene expression were established using fluorescence-based real-time PCR.[2, 22] Oligonucleotide primer pairs were generated using Primer3 software

(www.genome.wi.mit.edu), designed to generate 200 ± 1 base pair PCR products at the same annealing temperature (see *Supplementary Materials*). mRNA levels were quantified using real-time fluorometric analysis, and relative mRNA abundance was determined from triplicate assays using the comparative threshold cycle number (Δ-Ct method).[23]

C. Chip design and manufacture. Gene expression profiles were determined using an oligonucleotide chip uniquely designed and fabricated for comparative analysis of platelet-expressed genes. The gene list was generated using microarray profiles from a cohort of normal (N=5) and ET (N=6) platelet mRNAs hybridized to the Affymetrix HU133A GeneChip;[2, 22] leukocyte RNA from three normal patients were used to delineate leukocyte gene expression profiles. Finalized, custom spotted microarrays contained 432 platelet-expressed and 43 leukocyte-restricted genes which co-segregated by cell-type (platelet *vs.* leukocyte).[24] *Arabidopsis* probe elements were included for normalization controls and as quantitative measures of inter- and intra-slide variability;[25] 70-mer oligonucleotides were synthesized based on the Ensemble (www.ensembl.org/) Human 13.31 Database; all probe-sets were spotted in quadruplicate to provide replicates and statistical robustness.

D. Gene expression analysis. Platelet gene profiling was completed using a template-switching mechanism to optimize amplification from low-abundance mRNAs.[26] Initially, 20 ng of purified platelet or human reference RNA (Stratagene) was supplemented with a fixed amount of *Arabidopsis* mRNA to provide internal standards for hybridization and normalization. Chimeric DNA/RNA amplification and labeling was completed using the Ovation Aminoallyl system (NuGen Technologies), providing for 4 - 6 μg of cDNA/sample. cDNA solutions were vacuum-dried and coupled to Cy3 (human reference RNA) or Cy5 (patient RNA) dyes (Amersham Biosciences), and stoichiometrically equivalent mixes were hybridized to platelet chips prior to

gene quantification using a Gene Pix 4000B scanner (Molecular Devices). All microarray data were submitted to the GEO database in MIAME-compliant form, reported under NCBI accession #15472151. Initial data processing (gridding, technical spot analysis, etc.) was completed using GenePix Pro software. After rigorous inspection to exclude spotting irregularities, raw Cy3:Cy5 ratios were quantified for individual genes. Reproducibility of microarray profiles using biological replicates from healthy donors was excellent (Spearman correlation coefficients of 0.93 - 0.95).

E. Bioinformatics and statistical analyses. Microarray data were analyzed and visualized using GeneSpring (Silicon Genetics) or in-house software. Expression data were sequentially normalized by spot, by gene, and by chip essentially as previously described,[2, 22] followed by a moderate filtering step to maximize our ability to identify differentially-expressed genes. Genes with fluorescence intensities <10 in more than 70% of probes were excluded from further analysis. For each gene, the four ratios were averaged and $\log_2$-transformed prior to data analysis. The Kruskal-Wallis, non-parametric one-way ANOVA was performed to identify differentially-expressed genes among the three cohorts (i.e. ET, RT, normal). The nonparametric Wilcoxon rank-sum test was used to examine median differences between two independent samples. This included gender effects, the comparison between ET and RT subjects, as well as comparison within ET subjects by Jak2 genotype using either microarray or qRT-PCR data. The significance level is set at 0.05 (two-sided) unless otherwise specified.

Stepwise discriminant analysis was used to identify an initial biomarker subset that separated class on the basis of microarray data. The fidelity of the genetic biomarker subsets as class prediction tools was established using non-parametric linear discriminant analysis (LDA) with a leave-one-out cross-validation analysis.[27] An integral property of the leave-one-out cross-

validation estimator is that it provides an almost unbiased estimate of the generalizability of the genetic classifier. Posterior classification probability for each subject was derived and the binary decision was made for group assignment based on subject highest probability. As part of the confirmatory studies, the same biomarker set using the microarray data was applied to the qRT-PCR data, and fidelity established using LDA leave-one-out cross-validation. This same biomarker identification and validation procedure was applied both for separation of ET *vs*. RT, and for substratification of ET by Jak2 genotype (Jak2V^{617}F vs. wild-type alleles).

RESULTS

A. Patient cohort analysis. A total of 95 patients (ET [N=24]; RT [N=23]; healthy controls [N=48]) were recruited into the study (**Table 1**). The mean platelet counts for ET and RT patient cohorts were nearly identical and not statistically different. At the time of platelet collection, 4/24 ET and 1/23 RT patients had normal platelet counts, reflecting either medication (ET) or thrombocytotic resolution (RT). Compared to RT patients, a greater percentage of ET patients were ≥ 60 years of age or were female, reflecting known demographics of the disease.[20] Of the ET patients, 46% were heterozygous for the Jak2V^{617}F mutant (GT) allele, while a smaller fraction (8%) were found to be homozygous for the mutation (TT); these characteristics are consistent with those previously reported.[12, 14] No RT or healthy control patients harbored the Jak2V^{617}F mutation.

B. Gender effect on gene expression profiles. Given previous evidence demonstrating genetic differences between normal and ET platelets,[22] a platelet-focused gene chip for screening larger patient cohorts was fabricated. Initially, we sought to exclude any gender effect among the three cohorts, and calculated the Wilcoxon rank-sum test for each of the 423 genes on the array. For both normal and ET cohorts, the preponderance of the genes were equally distributed within the 95% confidence interval (CI), with only four of the genes in either group demonstrating any gender effect (**Fig. 1,** *Panel A*). In normal patients, two genes displayed greater expression in males (*MBOAT2* and *H2BF*), while two other genes were differentially-weighted towards females (*LOC152719* and *LOC390354*). In ET platelets, a single gene (*E2F1*) was male-skewed, while three genes (*GAS2L1*, *CXorf9*, and *PPME1*) were female-biased. In contrast, gender effects were more prominent in the RT cohort, with 12 genes falling outside the 95% CI, all of which demonstrated male-skewed gene expression differences. Two of these 12 genes (*ITGA2B*

and *ITGB3*) encode the major polypeptide subunits of the platelet glycoprotein IIB/IIIA (α_{IIb}/β_{III}) integrin, suggesting that the molecular mechanisms that control gene expression of the heterodimeric receptor complex are concordantly regulated by gender during situations associated with reactive thrombocytosis.

C. Delineation of a genetic biomarker subset for discriminant analysis. Of the genes on the microarray chip, 267 displayed expression values that were significantly different among the three groups ($p < 0.05$), as established using the Kruskal-Wallis non-parametric one-way ANOVA. Among this subset, 148 genes were found to be significantly different between RT and ET cohorts using the Wilcoxon rank sum test. Stepwise LDA identified an 11-biomarker subset that segregated the three phenotypic cohorts (ET vs. RT vs. normal) (**Fig 1**, *Panel B*). The utility of the initial 11-biomarker subset to predict class was confirmed using a leave-one-out cross-validation analysis, in which each case is classified by the profiles derived from all cases excluding that case.[27] This approach confirmed the generalizability of the statistical classifier (i.e. its performance on previously unseen data) by using the available data as both training and test data, thereby providing an unbiased estimate of class prediction. The posterior classification probabilities applied in a binary decision model using this gene-set for 3-cohort analysis confirmed that 82/95 (86.3%) of all patients could be correctly classified, with the most accurate prediction evident for RT patients (100%) (**Fig. 1**, *Panels C* and *D*). Of the patients with thrombocytosis, 1/47 (2.1%) was genetically classified as normals, although 6/48 (12.5%) normals were incorrectly classified as belonging to ET (N = 1) or RT (N = 5) cohorts. Gene Ontology annotations using GOstat[28] demonstrated that the predominant functional annotations of this 11-biomarker subset were related to nucleosome and chromatin assembly (genes *H3FA*,

HIST1H2AG, *WASF3*), critically important components of gene transcription that may regulate megakaryocytopoeisis and/or proplatelet formation.[29]

D. Thrombocytosis classification and prediction models. Since a primary goal of classification is to predict phenotypic class among patients with thrombocytosis, we used the 11-biomarker subset as a discriminatory tool restricted to ET and RT. Two-cohort LDA confirmed that ET and RT profiles segregated by class, with only 3/47 (6.4%) outliers and an overall classification success rate of 93.6% (**Fig. 2,** *Panel A* and **Table 2**). Indeed, the ET *vs.* RT discriminatory analyses were more accurate using a 2-cohort (as opposed to the 3-cohort) linear discriminatory classifier. Similar to the results evident in the 3-cohort analysis, RT was correctly classified in 100% of cases using microarray analysis, although 3/24 ET patients were classified as having RT. Of the 3 misclassified ET patients, two were $Jak2V^{617}F$-negative and ET was diagnosed by exclusion. Interestingly, while the third patient was heterozygous for the $Jak2V^{617}F$ mutant allele, an advanced stage cervical carcinoma was subsequently diagnosed within 8 months of the ET diagnosis. Since malignancy is known to cause thrombocytosis, this case is especially informative since it confirms the dynamic state of the platelet transcriptome, and implies that an ET profile can become "RT-like" in the presence of an appropriate stimulus.

As an additional validation for fidelity of the 11-member gene subset to discriminate between ET and RT cohorts, platelet gene profiles were re-analyzed using a distinct, confirmatory platform. We generated oligonucleotide primers to the 11-biomarker gene set, and completed quantitative RT-PCR (qRT-PCR) for a randomly selected subset of 10 patients in each cohort (see *Supplementary Materials).* Six of the biomarkers were found to have significantly different median expression levels between ET and RT cohorts *via* qRT-PCR at $p < 0.05$ (*CTNS*, *NGFRAP1*, *CLEC1B*, *H3F3A*, *APP*, and *TMP1*) (**Fig. 2**, *Panel B*). These

confirmatory results provided strong presumptive evidence that ET and RT profiles were genetically distinct using two independent platforms. Finally, the ability to assign class using either genetic platform was compared using LDA with leave-one-out cross-validation analysis. As shown in **Table 2**, binary class prediction using either microarray or qRT-PCR data alone gave remarkably similar and highly accurate phenotypic classification, with >90% accuracy using either approach [93.6% success using microarray profiles and 90% using qRT-PCR].

E. Substratification by Jak2V^{617}F genotype. We applied the identical discriminant and validation analyses for ET class prediction sub-stratified by the Jak2V^{617}F allele. Stepwise discriminant analysis based on the microarray data alone resulted in a 4-member subset comprised of genes *HIST1*, *SRP72*, *C20orf103*, and *CRYM*. LDA with cross-validation based on the microarray data alone confirmed that 20/23 (87%) patients were correctly classified, with the best classification evident in patients lacking the Jak2V^{6717}F mutant allele (12/12 patients properly classified). Comparable results were seen using qRT-PCR (as a validation platform) in which only 1/12 Jak2-wild type patients was misclassified (**Table 3**). In contrast, class prediction was less robust in the subset of patients harboring the Jak2V^{6717}F allele. By microarray, 3/11 Jak2V^{6717}F patients were misclassified, while qRT-PCR resulted in a higher misclassification rate (5/11). Thus, although the overall classification using confirmatory qRT-PCR remained satisfactory (74% correct classification), the results were not as robust as those using microarray. For both approaches, the specificity of the classification was best in the subset of patients that were Jak2 homozygous normal.

DISCUSSION

We now provide evidence that gene expression profiles can be used to classify platelet phenotypes using routine phlebotomy, with clinical implications for patients with thrombocytosis. The proper classification of thrombocytosis is of paramount importance given the frequency of thrombohemorrhagic complications known to occur in ET (to the exclusion of RT).[7, 8, 20] While the presence of the Jak2V^{617}F mutation is strong presumptive evidence for the diagnosis of ET, the absence of the mutation in up to 60% of ET patients highlights its limited applicability in the larger cohort of thrombocytotic patients. By applying a novel microarray platform to a comprehensive patient cohort, we have identified an 11-member gene biomarker subset that effectively predicts phenotypic class in 86% of aggregate platelet samples, with >90% classification success when specifically applied to patients with thrombocytosis. Highly-robust class prediction models for thrombocytosis were established using both microarray profiling and independently confirmed by platelet PCR, serving to confirm the validity of these conclusions through two autonomous platforms. While comparable approaches have previously been developed for patients with various human malignancies,[30, 31] these data now provide compelling evidence that transcriptomic approaches can be applied to platelet disorders.

Of the 11-biomarker subset initially identified by microarray profiling, 6 genes demonstrated statistically-significant gene expression differences between ET and RT using qRT-PCR. Of these genes, *APP* (encoding the amyloid β-precursor protein, AβPP) remains the best-characterized in platelets, with limited platelet functional information on *NGFRAP1*, *CLEC1B*, *CTNS*, *H3F3A*, or *TPM1*. AβPP is the precursor to the amyloid β-peptide (Aβ) that accumulates in the brains of patients with Alzheimer's disease. AβPP is an abundant platelet α-granule protein that is released upon platelet activation, and contains a Kunitz-type serine

protease inhibitor domain analogous to protease nexin-2 (PN2).[32] PN2/AβPP is a potent inhibitor of various serine proteases, notably coagulation factors IXa, XIa, Xa, and the tissue factor-FVIIa complex. Megakaryocyte/platelet overexpression of human AβPP in transgenic mice modulates cerebral thrombotic risk, suggesting that changes in platelet AβPP expression may affect the thrombohemorrhagic balance *in vivo*.[33] While our studies were not focused on delineating genetic differences that could affect platelet-related thrombosis or hemorrhage known to occur in ET patients, these collective observations implicate AβPP as a thrombohemorrhagic biomarker. Comparable gene profiling studies focusing on platelet transcriptomic differences stratified by ET thrombohemorrhagic events may provide greater insight into this issue.

Limited information exists on gender-restricted genetic differences in platelets, despite evidence for differences in platelet function or platelet-targeted therapies.[34, 35] Our preliminary studies identified gene subsets that may be gender-skewed in normal and ET platelets, although these differences are more pronounced in RT patients. In this cohort, a male-weighted trend in higher gene expression was evident for 12 genes. Unlike the situation with ET or normal platelets, none of the RT gene expressions were weighted towards females. While these observations are preliminary, they are noteworthy since two genes (*ITGA2B* and *ITGB3*) displayed congruent expression changes, and encode both polypeptide chains of the GPIIb/IIIa fibrinogen receptor, the final common mediator of platelet aggregation. Recent meta-analysis on efficacy of GPIIb/IIIa inhibitors in acute coronary syndromes confirmed reduced risk of death or myocardial infarction in men, but increased risk in women.[36] Given the link between inflammation and thrombosis,[37] it is plausible that gender-related *ITGA2B/ITGB3* gene

expression changes could be relevant in the context of localized inflammatory responses present within the coronary vasculature.

Despite evidence that RT and ET (and normal) platelets are genetically distinct, analysis of larger patient cohorts will be required to determine if the various RT subtypes can be further subclassified (i.e. malignancy-associated, inflammatory, iron-deficiency, etc.). In inflammation-associated thrombocytosis, interleukin-6 (IL-6) may function as a critical mediator by upregulating thrombopoietin mRNA.[38] While a comparable IL-6 role may be evident in iron-deficiency or malignancy-associated thrombocytosis,[7] it remains likely that additional mechanisms may be operational.[39] Since many of these ligands mediate their effects through receptors known to regulate transcription, we speculate that expression profiles may be further substratified by etiology, a hypothesis readily testable with more extensive cohort analysis.

Although our data demonstrated robust classification success between ET and RT cohorts, substratification of ET by Jak2 genotype was less productive. A 4-biomarker gene subset effectively classified ET containing the Jak2 GG (normal) genotype using either microarray (100%) or qRT-PCR (92%), although class prediction was less efficient in patients harboring Jak2V^{617}F allele(s). One explanation may be the limited number of evaluable patients used for the discriminatory classifier. Alternatively, the transcript profiles of Jak2V^{617}F platelets may be heterogeneous, with both phenotypic and genotypic differences affected by quantitative burden of mutant JakV^{617}F.[40]

FIGURE LEGENDS

FIGURE 1. *Gender and discriminant analysis based on three-cohort microarray expression profiles.*

Panel A. Scatter plots by patient cohort were generated by applying a non-parametric Wilcoxon ranked-sum test to determine gender differences in gene expression for each of the 432 genes on the microarray chip. The z-score represents a weighted difference between the observed and the expected rank sums for the Wilcoxon test; thus, a z-score of 0 represents equivalent median expression (male gene expression = female gene expression), while the dashed lines represent 95% confidence intervals that are female (positive)- or male (negative)-skewed. The individual genes that are differentially expressed by gender are delineated in each phenotypic group. RT – Reactive thrombocytosis; ET – Essential Thrombocytosis

Panel B. The 11-biomarker subset identified by discriminant analysis is displayed by gene name. P-values from the three-cohort (ET *vs.* RT. *vs.* normal) Kruskal-Wallis non-parametric ANOVAs are displayed by individual gene.

Panel C. Plot of posterior classification probability demonstrates the segregation of the three phenotypes (ET = 24, RT = 23; Normal = 48) using the 11-biomarker gene subset *via* linear discriminant analysis with leave-one-out cross validation (group centroids are depicted). Each symbol represents one patient microarray that incorporates aggregate expression data from the 11-gene subset. For each patient sample, the posterior classification probabilities that a given subject belongs to ET or RT cohorts are displayed by subject index (the probability that a subject belongs to the normal cohort is equivalent to 1 minus the sum of the corresponding ET and RT probabilities).

Panel D. Phenotypic binary class prediction using the same algorithm and 11-biomarker gene subset (numbers in parentheses are percent by group). A subject is classified to the group corresponding to his/her highest posterior classification probability.

FIGURE 2. *Thrombocytosis discriminant analysis using an 11-biomarker gene subset.*

Panel A. Linear discriminant analysis plot shows the posterior classification probability of each subject by cohort (RT = 23 and ET = 24), using the 11-biomarker gene subset based on microarray profiles (group centroids are shown). For binary decisions, a subject is classified based on the highest posterior classification (i.e. a group with probability >0.5). Patient outliers are delineated by arrows.

Panel B. A randomly-selected subset of ET (N = 10) and RT (N = 10) patient platelets were analyzed by qRT-PCR using oligonucleotides primers specific to each of the 11 genetic biomarkers. Relative gene expression is displayed on a $\log_{10}$ scale standardized to β-actin mRNA, calculated from triplicate wells for each patient sample using the comparative threshold number (Δ−Ct). Boxes represent the interquartile range that encompasses 50% of the values, while the 95% confidence intervals and outliers are depicted; the horizontal bar within each box marks the group median (p-values were calculated using non-parametric Wilcoxon rank sum test).

ACKNOWLEDGEMENTS

We thank Xiao Wu for statistical analysis and Lesley Scudder and Jean Wainer for excellent technical support. This research was supported by grants HL49141 and HL086376 (National Institutes of Health), MP048005 (Department of Defense), NIH Center Grant MOI 10710-5 to the Stony Brook University General Clinical Research Center, and a Targeted Research Award (Stony Brook University).

REFERENCES

1. Bahou WF. Genetic disorders of platelets. In: Kumar D, ed. Genomics and Clinical Medicine. Oxford: Oxford Univesity Press; 2006.
2. Gnatenko DV, Dunn JJ, McCorkle SR, Weissmann D, Perrotta PL, Bahou WF. Transcript profiling of human platelets using microarray and serial analysis of gene expression. Blood 2003;101(6):2285-93.
3. McRedmond JP, Park SD, Reilly DF, et al. Integration of Proteomics and Genomics in Platelets: A profile of platelet proteins and platelet-specific genes. Mol Cell Proteomics 2004;3(2):133-44.
4. Denis MM, Tolley ND, Bunting M, et al. Escaping the nuclear confines: signal-dependent pre-mRNA splicing in anucleate platelets. Cell 2005;122(3):379-91.
5. Weyrich A, Dixon D, Pabla R, et al. Signal-dependent translation of a regulatory protein, Bcl-2, in activated human platelets. ProcNatAcadSciUSA 1998;95:5556-61.
6. Rinder H, Schuster J, Rinder C, Wang C, Schwseidler H, Smith B. Correlation of thrombosis with increased platelet turnover in thrombocytosis. Blood 1998;91:1288-94.
7. Schafer AI. Thrombocytosis. The New England Journal of Medicine 2004;350(12):1211-9.
8. Griesshammer M, Bangerter M, Sauer T, Wennauer R, Bergmann L, Heimpel H. Aetiology and clinical significance of thrombocytosis: analysis of 732 patients with an elevated platelet count. Journal of Internal Medicine 1999;245(3):295-300.
9. Bazzan M, Tamponi G, Schinco P, et al. Thrombosis-free survival and life expectancy in 187 consecutive patients with essential thrombocythemia. Ann Hematol 1999;78(12):539-43.
10. Zahavi J, Zahavi M, Firsteter E, Frish B, Turleanu R, Rachmani R. An abnormal pattern of multiple platelet function abnormalities and increased thromboxane generation in patients with primary thrombocytosis and thrombotic complications. Eur J Haematol 1991;47(5):326-32.
11. Shih LY, Lin TL, Lai CL, et al. Predictive values of X-chromosome inactivation patterns and clinicohematologic parameters for vascular complications in female patients with essential thrombocythemia. Blood 2002;100(5):1596-601.
12. Baxter EJ, Scott LM, Campbell PJ, et al. Acquired mutation of the tyrosine kinase JAK2 in human myeloproliferative disorders. Lancet 2005;365(9464):1054-61.
13. James C, Ugo V, Le Couedic JP, et al. A unique clonal JAK2 mutation leading to constitutive signalling causes polycythaemia vera. Nature 2005;434(7037):1144-8.
14. Kralovics R, Passamonti F, Buser AS, et al. A gain-of-function mutation of JAK2 in myeloproliferative disorders. The New England Journal of Medicine 2005 Apr 28:1779-90.
15. Finazzi G, Harrison C. Essential thrombocythemia. Seminars in Hematology 2005;42(4):230-8.
16. Bhattacharjee A, Richards WG, Staunton J, et al. Classification of human lung carcinomas by mRNA expression profiling reveals distinct adenocarcinoma subclasses. Proceedings of the National Academy of Sciences of the United States of America 2001;98(24):13790-5.
17. Shipp MA, Ross KN, Tamayo P, et al. Diffuse large B-cell lymphoma outcome prediction by gene-expression profiling and supervised machine learning. Nat Med 2002;8(1):68-74.
18. Ramaswamy S, Ross KN, Lander ES, Golub TR. A molecular signature of metastasis in primary solid tumors. Nature Genetics 2003;33(1):49-54.

19. Iland HJ, Laszlo J, Case DC, Jr., et al. Differentiation between essential thrombocythemia and polycythemia vera with marked thrombocytosis. Am J Hematol 1987;25(2):191-201.
20. Murphy S, Peterson P, Iland H, Laszlo J. Experience of the Polycythemia Vera Study Group with essential thrombocythemia: a final report on diagnostic criteria, survival, and leukemic transition by treatment. Seminars in Hematology 1997;34(1):29-39.
21. Mirza H, Yatsula V, Bahou WF. The proteinase activated receptor-2 (PAR-2) mediates mitogenic responses in human vascular endothelial cells. Molecular characterization and evidence for functional coupling to the thrombin receptor. J Clin Invest 1996;97:1705-14.
22. Gnatenko DV, Cupit LD, Huang EC, Dhundale A, Perrotta PL, Bahou WF. Platelets express steroidogenic 17beta-hydroxysteroid dehydrogenases. Distinct profiles predict the essential thrombocythemic phenotype. Thrombosis and Haemostasis 2005;94(2):412-21.
23. Schmidt VA, Chiariello CS, Capilla E, Miller F, Bahou WF. Development of hepatocellular carcinoma in Iqgap2-deficient mice is IQGAP1 dependent. Molecular and Cellular Biology 2008;28(5):1489-502.
24. Gnatenko DV, Perrotta PL, Bahou WF. Proteomic approaches to dissect platelet function: Half the story. Blood 2006;108(13):3983-91.
25. Wang HY, Malek RL, Kwitek AE, et al. Assessing unmodified 70-mer oligonucleotide probe performance on glass-slide microarrays. Genome Biology 2003;4(1):R5.
26. Wang E, Miller L, Ohnmacht G, Liu E, Marincola F. High-fidelity mRNA amplification for gene profiling. Nature Biotechnology 2000;18:157-459.
27. Cawley GC, Talbot NL. Fast exact leave-one-out cross-validation of sparse least-squares support vector machines. Neural Netw 2004;17(10):1467-75.
28. Beissbarth T, Speed TP. GOstat: find statistically overrepresented Gene Ontologies within a group of genes. Bioinformatics (Oxford, England) 2004;20(9):1464-5.
29. Schulze H, Shivdasani RA. Molecular mechanisms of megakaryocyte differentiation. Semin Thromb Hemost 2004;30(4):389-98.
30. Hedenfalk I, Duggan D, Chen Y, et al. Gene-expression profiles in hereditary breast cancer. The New England Journal of Medicine 2001;344(8):539-48.
31. Valk PJ, Verhaak RG, Beijen MA, et al. Prognostically useful gene-expression profiles in acute myeloid leukemia. The New England Journal of Medicine 2004;350(16):1617-28.
32. Van Nostrand WE, Wagner SL, Suzuki M, et al. Protease nexin-II, a potent antichymotrypsin, shows identity to amyloid beta-protein precursor. Nature 1989;341(6242):546-9.
33. Xu F, Davis J, Miao J, et al. Protease nexin-2/amyloid beta-protein precursor limits cerebral thrombosis. Proceedings of the National Academy of Sciences of the United States of America 2005;102(50):18135-40.
34. Collaborative meta-analysis of randomised trials of antiplatelet therapy for prevention of death, myocardial infarction, and stroke in high risk patients. BMJ 2002;324(7329):71-86.
35. Johnson M, Ramey E, Ramwell PW. Sex and age differences in human platelet aggregation. Nature 1975;253(5490):355-7.
36. Boersma E, Harrington RA, Moliterno DJ, et al. Platelet glycoprotein IIb/IIIa inhibitors in acute coronary syndromes: a meta-analysis of all major randomised clinical trials. Lancet 2002;359(9302):189-98.
37. Esmon CT. Inflammation and thrombosis. J Thromb Haemost 2003;1(7):1343-8.
38. Kaser A, Brandacher G, Steurer W, et al. Interleukin-6 stimulates thrombopoiesis through thrombopoietin: role in inflammatory thrombocytosis. Blood 2001;98(9):2720-5.

39. Li AJ, Karlan BY. Androgen mediation of thrombocytosis in epithelial ovarian cancer biology. Clin Cancer Res 2005;11(22):8015-8.
40. Antonioli E, Guglielmelli P, Poli G, et al. Influence of JAK2V617F allele burden on phenotype in essential thrombocythemia. Haematologica 2008;93(1):41-8.

TABLE 1. PATIENT CHARACTERISTICS

Characteristic	Essential Thrombocytosis[†] (N = 24)	Reactive Thrombocytosis (N = 23)	Normal controls[‡] (N = 48)
Gender – Number (%)			
Male	8 (33)	12 (52)	23 (48)
Female	16 (67)	11 (48)	25 (52)
Age group – Number (%)			
<35	3 (12)	5 (22)	16 (33)
36-60	6 (25)	15 (65)	27 (57)
≥ 60	15 (63)	3 (13)	5 (10)
Platelet count [$x10^{-3}/mm^3$]			
Mean	703	675	246
Range	220 - 1724	173 – 1044	160 -415
Hemoglobin [g/dL]			
Mean	13.3	11.1	14.1
Range	9.5 – 20.8	8.1 – 15.3	12.3 – 15.9
White cell count [$x10^{-3}/mm^3$]			
Mean	10.5	16.2	6.3
Range	4.3 – 44.6	6.9 – 65.2	4.4 – 9.6
Jak2 Genotype – No. (%)			
$V^{617}F$ Heterozygous	9 (38)	0 (0)	0 (0)
$V^{617}F$ Homozygous	2 (8)	0 (0)	0 (0)
Normal	13 (54)	23 (100)	48 (100)

[†]At the time of platelet isolation, 10 patients were being treated with hydroxyurea, 8 patients with anagrelide, and 6 were untreated
[‡]Random subset of 10 males and 10 females

TABLE 2. GENETIC CLASSIFICATION OF THROMBOCYTOSIS USING AN 11-BIOMARKER GENE SUBSET

PHENOTYPIC CLASS	GENOTYPIC CLASS[†]					
	Microarray profiles			**qRT-PCR**		
	ET	RT	**Total**	ET	RT	**Total**
ET	21 (87.5)[‡]	3 (12.5)	24	9 (90.0)	1 (10.0)	10
RT	0 (0)	23 (100)	23	1 (10.0)	9 (90.0)	10
Total	21	26	47	10	10	20
Classification by group aggregate			44/47 (93.6)			18/20 (90.0)

[†]Classification was completed using non-parametric linear discriminant analysis with leave-one-out cross-validation

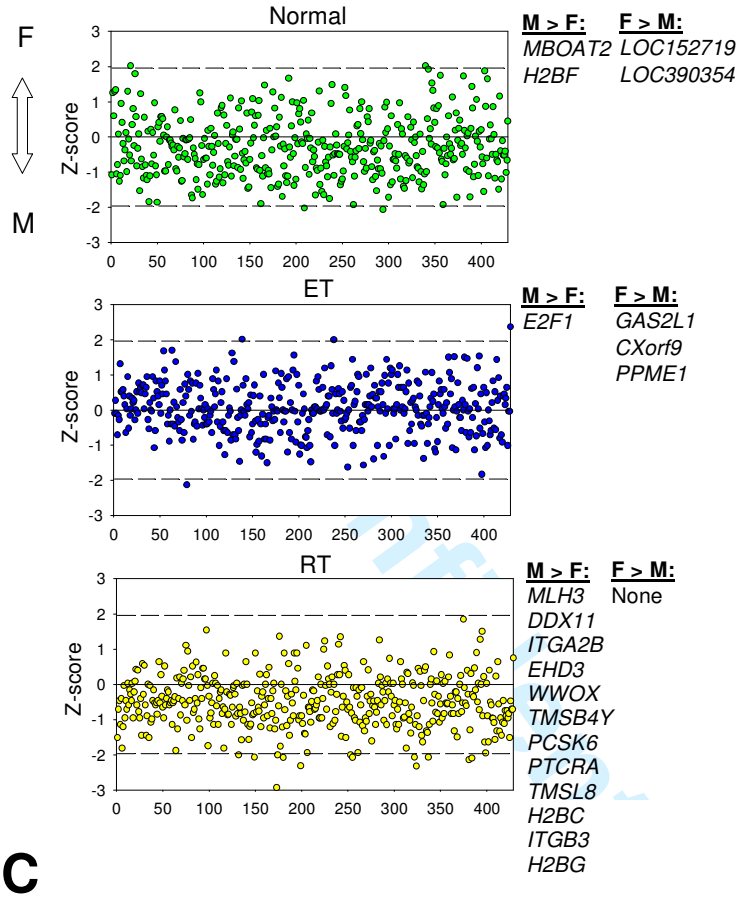
[‡]Numbers in parentheses are percent

TABLE 3. JAK2 CLASS PREDICTION AND DISCRIMINANT ANALYSIS FUNCTIONS USING A FOUR-BIOMARKER GENE SUBSET

	CLASS PREDICTION						GENE DISCRIMINANT POWER[†]				
Genotype[‡]	**Microarray**			**qRT-PCR**				**Microarray**		**qRT-PCR**	
	Jak2 (GG)	Jak2 (GT/TT)	Total	Jak2 (GG)	Jak2 (GT/TT)	Total	Gene[§]	Jak2 (GG)	Jak2 (GT/TT)	Jak2 (GG)	Jak2 (GT/TT)
Jak2 (GG)	12 (100.0)	0 (0)	12	11 (91.7)	1 (8.3)	12	*HIST1H1A*	91.6	36.4	100	27.3
Jak2 (GT/TT)	3 (27.3)	8 (72.7)	11	5 (45.5)	6 (54.6)	11	*SRP72*	91.6	36.4	33.3	62.6
							C20orf103	16.7	100	16.7	90.9
Group Aggregates			20/23 (87.0)			17/23 (73.9)	*CRYM*	75	54.6	91.7	9.1

[†]Discriminant power of each individual gene was determined by entering only the given gene in the linear discriminant analysis
[‡]GG (wild-type normal); GT (Jak2V^{617}F heterozygote); TT (Jak2V^{617}F homozygote)
[§]*HIST1H1A* - Histone Cluster 1, H1a*; SRP72* – Signal recognition particle 72kDa; *C20orf103* – Chromosome 20 open reading frame 103; *CRYM* – Crystalline, mu

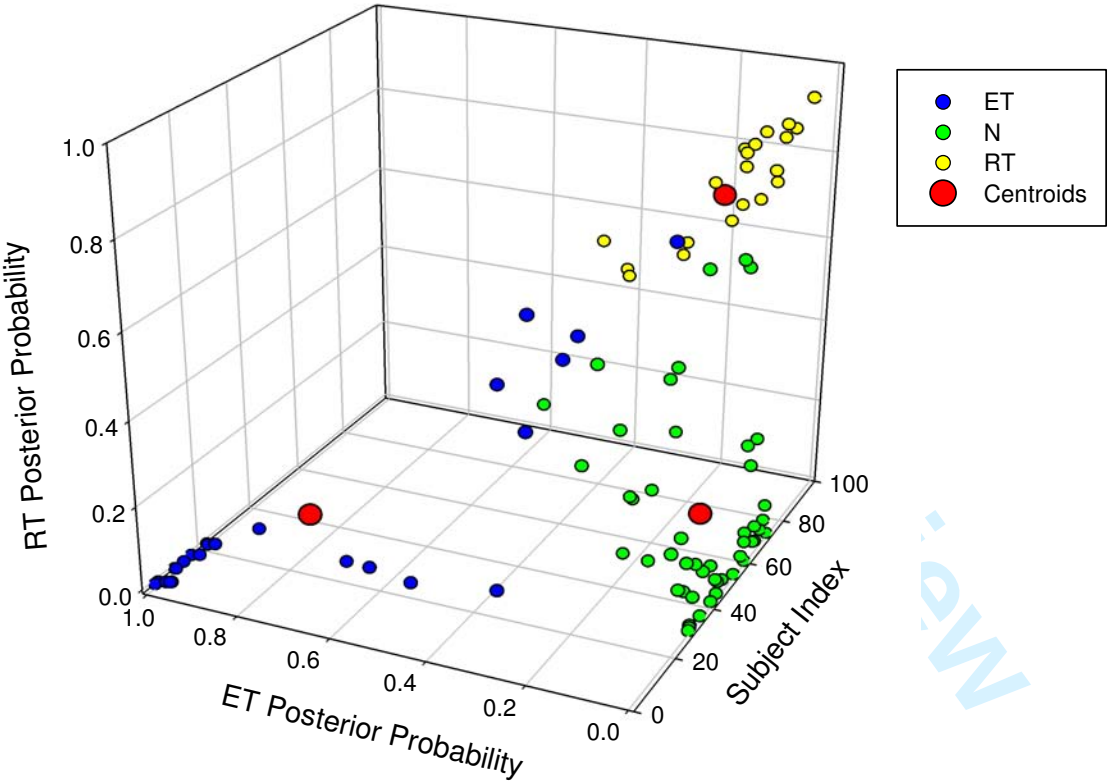
A



B

Gene name	Description	p-value
WASF3	WAS protein family, member 3	< 0.001
CTNS	Cystinosis, nephropathic	< 0.001
HIST1H2AG	Histone 1, H2ag	< 0.001
ACOT7	Acyl-CoA thioesterase 7	0.13
LAPTM4B	Lysosomal associated protein	0.03
TGFB2	Transforming growth factor-β precursor	< 0.001
TPM1	Tropomyosin	0.26
H3F3A	H3 histone, family 3B	< 0.001
APP	Amyloid-β (A4) precursor protein	0.03
NGFRAP1	NGF receptor associated protein 1	0.07
CLEC1B	C-type lectin domain family 1, memb *B*	< 0.001

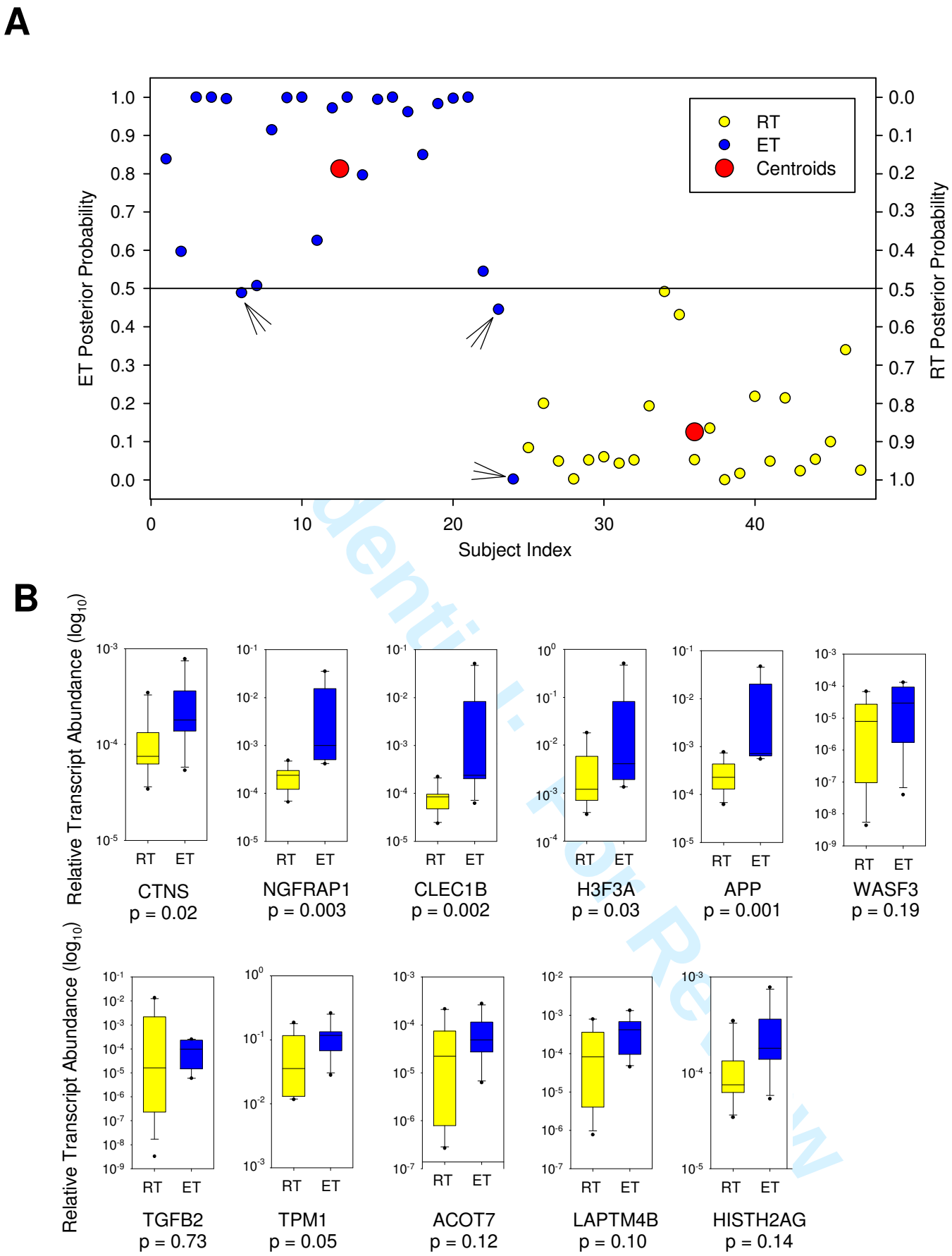
C



D

	Normal	RT	ET	Total
Normal	42 (87.5)	5 (10.4)	1 (2.1)	48
RT	0	23 (100)	0	23
ET	1 (4.2)	6 (25.0)	17 (70.8)	24
Classification by group aggregate				**82/95 (86.3)**

Gnatenko, et. al., Figure 1



Gnatenko, et. al., Figure 2